

Phi-divergence statistics for the likelihood ratio order: an approach based on log-linear models

Martin, N.¹, Mata, R.² and Pardo, L.^{2*}

¹Department of Statistics, Carlos III University of Madrid, 28903 Getafe (Madrid), Spain

²Department of Statistics and O.R., Complutense University of Madrid, 28040 Madrid, Spain

February 24, 2014

Abstract

When some treatments are ordered according to the categories of an ordinal categorical variable (e.g., extent of side effects) in a monotone order, one might be interested in knowing whether the treatments are equally effective or not. One way to do that is to test if the likelihood ratio order is strictly verified. A method based on log-linear models is derived to make statistical inference and phi-divergence test-statistics are proposed for the test of interest. Focussed on loglinear modeling, the theory associated with the asymptotic distribution of the phi-divergence test-statistics is developed. An illustrative example motivates the procedure and a simulation study for small and moderate sample sizes shows that it is possible to find phi-divergence test-statistic with an exact size closer to nominal size and higher power in comparison with the classical likelihood ratio.

Keywords and phrases: Phi-divergence test statistics, Inequality constraints, Likelihood ratio ordering, Loglinear modeling.

1 Introduction

In this paper we are interested in comparing I treatments when the response variable is ordinal with J categories. We can consider each treatment type to be each of the I ordinal categories of a variable X . We shall denote by Y the response variable and its conditional probabilities by

$$\boldsymbol{\pi}_i = (\pi_{i1}, \dots, \pi_{iJ})^T, \quad i = 1, \dots, I,$$

with

$$\pi_{ij} = \Pr(Y = j | X = i), \quad j = 1, \dots, J.$$

For the i -th treatment and for each individual taken independently from a sample of size n_i its response is classified to be $\{1, \dots, J\}$ according to the conditional distribution of $Y | X = i$, $\boldsymbol{\pi}_i$. In this setting the J -dimensional random variable associated with the observed frequencies,

$$\mathbf{N}_i = (N_{i1}, \dots, N_{iJ})^T,$$

is multinomially distributed with parameters n_i and $\boldsymbol{\pi}_i$. Assuming that the different treatments are independent, the probability distribution of the $I \times J$ dimensional random variable $\mathbf{N} = (\mathbf{N}_1^T, \dots, \mathbf{N}_I^T)^T$ is product-multinomial. We are going to consider a motivation example, taken from Section 5 in Dardanoni and Forcina

*Corresponding author. E-mail: lpardo@mat.ucm.es, Tel: (+34) 91 394 4425, Fax: (+34) 91 394 4606.

(1998), in order to clarify the problem considered in this paper. In Table 1, duodenal ulcer patients of a hospital are cross-classified according to an increasing order of $I = 4$ severity degrees of the operation, and the extent of side effects, categorized as None, Slight and Moderate ($J = 3$).

	None	Slight	Moderate
Treatment 1	61	28	7
Treatment 2	68	23	13
Treatment 3	58	40	12
Treatment 4	53	38	16

Table 1: Extent of size effect of four treatments.

We shall consider that Treatment $i + 1$ is as good as Treatment i , for $i = 1, \dots, I - 1$ simultaneously, if $\frac{\Pr(Y=j|X=i+1)}{\Pr(Y=j|X=i)}$ is non-decreasing for all $j \in \{1, \dots, J\}$, i.e.

$$\frac{\Pr(Y=j|X=i+1)}{\Pr(Y=j|X=i)} \leq \frac{\Pr(Y=j+1|X=i+1)}{\Pr(Y=j+1|X=i)}, \quad \text{for every } (i, j) \in \{1, \dots, I - 1\} \times \{1, \dots, J - 1\}. \quad (1)$$

This is the so called “likelihood ratio ordering”, sometimes called also “local ordering” (see Silvapulle and Sen (2005), Chapter 6). It is very important to clarify that the likelihood ratio ordering is more thoroughly referred to J independent multinomial samples of sizes equal to 1, in such a way that $\pi_{ij} = \Pr(Y = j|X = i) = \Pr(\mathcal{M}(1, \boldsymbol{\pi}_i) = \mathbf{e}_j)$, where \mathbf{e}_j is the j -th unit vector. In a similar way, Treatment $i + 1$ is better than Treatment i , for $i = 1, \dots, I - 1$ simultaneously, if (1) holds with at least one strict inequality. For testing that Treatment $i + 1$ is better than Treatment i , for $i = 1, \dots, I - 1$ we can consider

$$H_0 : \frac{\Pr(Y=j|X=i+1)}{\Pr(Y=j|X=i)} = \frac{\Pr(Y=j+1|X=i+1)}{\Pr(Y=j+1|X=i)} \quad \text{for every } (i, j) \in \{1, \dots, I - 1\} \times \{1, \dots, J - 1\}, \quad (2a)$$

$$H_1 : \frac{\Pr(Y=j|X=i+1)}{\Pr(Y=j|X=i)} \leq \frac{\Pr(Y=j+1|X=i+1)}{\Pr(Y=j+1|X=i)} \quad \text{for every } (i, j) \in \{1, \dots, I - 1\} \times \{1, \dots, J - 1\} \quad (2b)$$

$$\text{and } \frac{\Pr(Y=j|X=i+1)}{\Pr(Y=j|X=i)} < \frac{\Pr(Y=j+1|X=i+1)}{\Pr(Y=j+1|X=i)} \quad \text{for at least one } (i, j) \in \{1, \dots, I - 1\} \times \{1, \dots, J - 1\}.$$

For the motivation example, the null hypothesis means that all the treatments have equal side effects, while the alternative hypothesis means that as the more severe treatment is, greater is the probability of having side effects. Note that if we multiply on the left and right hand side of (2a) and (2b) by $\left(\frac{\Pr(Y=j|X=i+1)}{\Pr(Y=j|X=i)}\right)^{-1}$ we obtain

$$H_0 : \vartheta_{ij} = 1, \quad \forall (i, j) \in \{1, \dots, I - 1\} \times \{1, \dots, J - 1\}, \quad (3a)$$

$$H_1 : \vartheta_{ij} \geq 1 \quad \text{for every } (i, j) \in \{1, \dots, I - 1\} \times \{1, \dots, J - 1\} \quad (3b)$$

$$\text{and } \vartheta_{ij} > 1 \quad \text{for at least one } (i, j) \in \{1, \dots, I - 1\} \times \{1, \dots, J - 1\},$$

where $\vartheta_{ij} = \frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i+1,j}\pi_{i,j+1}}$ represent the “local odds ratios”, also called cross-product ratios.

If we denote by $n = \sum_{i=1}^I n_i$ the total of the sample sizes, we can consider the joint distribution to be

$$p_{ij} = \Pr(X = i, Y = j) = \Pr(X = i) \Pr(Y = j|X = i) = \frac{n_i}{n} \pi_{ij}, \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

We display such a distribution in a rectangular table having I rows for the categories of X and J columns for the categories of Y , and we denote $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_I)^T$, with $\mathbf{p}_i = (p_{i1}, \dots, p_{iJ})^T$, $i = 1, \dots, I$, the corresponding $I \times J$ matrix and

$$\mathbf{p} = \text{vec}(\mathbf{P}^T) = (\mathbf{p}_1^T, \dots, \mathbf{p}_I^T)^T \quad (4)$$

a vector obtained by stacking the columns of \mathbf{P}^T (i.e., the rows of matrix \mathbf{P}). Note that the components of \mathbf{P} are ordered in lexicographical order in \mathbf{p} . The local odds ratios can be expressed only in terms of joint probabilities

$$\vartheta_{ij} = \frac{p_{ij}p_{i+1,j+1}}{p_{i+1,j}p_{i,j+1}} = \frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i+1,j}\pi_{i,j+1}}, \quad \forall (i, j) \in \{1, \dots, I-1\} \times \{1, \dots, J-1\}. \quad (5)$$

The likelihood ratio ordering has been extensively studied in order statistics. In the literature related to order restricted inference for categorical data analysis, the likelihood ratio ordering has received little attention. The definition given in (1) is not specific for multinomial random variables, actually is very similar for any random variable, not necessarily discrete. In Bapar and Kochar (1994), it is mentioned that very important families of random variables, such as the one-parameter exponential family of distributions, have the likelihood ratio ordering property with respect to the parameter. For two independent multinomial samples ($I = 2$), Dykstra et al. (1995) established the asymptotic distribution of the likelihood ratio test-statistic and Dardanoni and Forcina extended it for a general problem of I independent multinomial samples. Recently, Davidov et al. (2010) have highlighted its importance by considering it as a particular case of the power bias model. This ordering is important not only in fields such as order statistics and its consideration cannot be avoided in categorical data analysis. Dykstra et al. (1995) argued that the likelihood ratio ordering, applied in an adapted product-multinomial sampling context, is a useful method for making statistical inference related to trend comparison of Poisson processes. The merit of the work of Dardanoni and Forcina (1998) is not only in the results obtained for a more general case, but also in the parametrization used for the method developed to find the asymptotic distribution of the likelihood ratio-test. They studied three kinds of ordering in the same parametrization setting but any of them is considered to be superior, for example the likelihood ratio ordering is stronger in comparison with the stochastic one.

In Section 2 of this paper a new method is proposed, based on log-linear modeling, for characterizing the likelihood ratio test for the likelihood ratio ordering in several independent multinomial samples. None paper has considered any alternative test-statistic to the likelihood ratio one and hence it is interesting to study the performance of the phi-divergence test-statistics. These test-statistics, which include as a particular case the likelihood ratio one, are introduced in Section 3. It is proven, in Section 4, that the asymptotic distribution of the phi-divergence test-statistics is chi-bar and an algorithm is also provided to find its weights in a simple way. An illustrative example is given in Section 5, as well as an algorithm, to clarify the method and the computational aspects. We consider that the likelihood ratio ordering, as strong ordering type and nested model within other ordering types, is a useful order since it should have asymptotically, under an alternative hypothesis of likelihood ratio order, much power. One of our interests in this regard, is to study through simulation the performance of the likelihood ratio test-statistic for small and moderate multinomial sample sizes, as Wang (1996) did in relation to the stochastic order. This is done in Section 6.

2 Modeling local odds ratios through loglinear models

Our first aim in this paper is to formulate the hypothesis testing problem making a reparametrization using the saturated loglinear model associated with \mathbf{p} , so that the restrictions are linear with respect to the new parameters. Focussed on \mathbf{p} , the saturated loglinear model with canonical parametrization is defined by

$$\log p_{ij} = u + u_{1(i)} + \theta_{2(j)} + \theta_{12(ij)}, \quad (6)$$

with

$$u_{1(I)} = 0, \quad \theta_{2(J)} = 0, \quad \theta_{12(iJ)} = 0, i = 1, \dots, I-1, \quad \theta_{12(Ij)} = 0, j = 1, \dots, J. \quad (7)$$

It is important to clarify that we have used the identifiability constraints (7) in order to make easier the calculations. Similar conditions have been used for instance in Lang (1996, examples of Section 7) and Sen and Silvapulle (2005, exercise 6.25 in page 345). Let $\boldsymbol{\theta}_2 = (\theta_{2(1)}, \dots, \theta_{2(J-1)})^T$, $\boldsymbol{\theta}_{12(i)} = (\theta_{12(i1)}, \dots, \theta_{12(i, J-1)})^T$, $i = 1, \dots, I-1$, be the subvector of unknown parameters of , $\boldsymbol{\theta} = (\boldsymbol{\theta}_2^T, \boldsymbol{\theta}_{12(1)}^T, \dots, \boldsymbol{\theta}_{12(I-1)}^T)^T$ and the vector of

redundant components of the model $\mathbf{u} = (u, u_{1(1)}, \dots, u_{1(I-1)})^T$, since its components can be expressed in terms of $\boldsymbol{\theta}$ as follows

$$\begin{aligned} u &= u(\boldsymbol{\theta}) = \log n_I - \log n - \log \sum_{j=1}^J \exp\{\theta_{2(j)}\} \\ &= \log n_I - \log n - \log (1 + \mathbf{1}_{J-1}^T \exp\{\boldsymbol{\theta}_2\}), \end{aligned} \quad (8)$$

$$\begin{aligned} u_{1(i)} &= u_{1(i)}(\boldsymbol{\theta}) = \log n_i - \log n - \log \sum_{j=1}^J \exp\{\theta_{2(j)} + \theta_{12(ij)}\} - u(\boldsymbol{\theta}). \\ &= \log n_i - \log n_I + \log \left(1 + \sum_{j=1}^{J-1} \exp\{\theta_{2(j)}\}\right) - \log \left(1 + \sum_{j=1}^{J-1} \exp\{\theta_{2(j)} + \theta_{12(ij)}\}\right) \\ &= \log \frac{n_i}{n_I} + \log \frac{1 + \mathbf{1}_{J-1}^T \exp\{\boldsymbol{\theta}_2\}}{1 + \mathbf{1}_{J-1}^T \exp\{\boldsymbol{\theta}_2 + \boldsymbol{\theta}_{12(i)}\}}, \quad i = 1, \dots, I-1. \end{aligned} \quad (9)$$

Note that the expressions of $u_{1(i)} = u_{1(i)}(\boldsymbol{\theta})$, $i = 1, \dots, I-1$ and $u = u(\boldsymbol{\theta})$ are obtained taking into account that for product-multinomial sampling $\sum_{j=1}^J p_{ij}(\boldsymbol{\theta}) = \frac{n_i}{n}$, $i = 1, \dots, I$. In matrix notation, (6) is given by

$$\begin{aligned} \log \mathbf{p}(\boldsymbol{\theta}) &= \left(\begin{pmatrix} \mathbf{1}_{I-1} & \mathbf{I}_{I-1} \\ 1 & \mathbf{0}_{I-1}^T \end{pmatrix} \otimes \begin{pmatrix} \mathbf{1}_{J-1} & \mathbf{I}_{J-1} \\ 1 & \mathbf{0}_{J-1}^T \end{pmatrix} \right) \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\theta} \end{pmatrix} \\ &= \mathbf{W}_0 \mathbf{u} + \mathbf{W} \boldsymbol{\theta}, \end{aligned} \quad (10)$$

where \otimes is the Kronecker product (see Chapter 16 of Harville (2008)), \mathbf{I}_a is the identity matrix of order a , $\mathbf{0}_a$ is the a -vector of zeros, $\mathbf{p}(\boldsymbol{\theta})$ is \mathbf{p} such that the components are defined by (6) and

$$\mathbf{W}_0 = \begin{pmatrix} \mathbf{1}_{I-1} & \mathbf{I}_{I-1} \\ 1 & \mathbf{0}_{I-1}^T \end{pmatrix} \otimes \mathbf{1}_J, \quad \mathbf{W} = \begin{pmatrix} \mathbf{1}_{I-1} & \mathbf{I}_{I-1} \\ 1 & \mathbf{0}_{I-1}^T \end{pmatrix} \otimes \begin{pmatrix} \mathbf{I}_{J-1} \\ \mathbf{0}_{J-1}^T \end{pmatrix}.$$

Condition (1) can be expressed by the linear constraint

$$\theta_{12(ij)} - \theta_{12(i+1,j)} - \theta_{12(i,j+1)} + \theta_{12(i+1,j+1)} \geq 0, \quad \forall (i,j) \in \{1, \dots, I-1\} \times \{1, \dots, J-1\}, \quad (11)$$

because

$$\log \vartheta_{ij} = \log p_{ij} - \log p_{i+1,j} - \log p_{i,j+1} + \log p_{i+1,j+1} = \theta_{12(ij)} - \theta_{12(i+1,j)} - \theta_{12(i,j+1)} + \theta_{12(i+1,j+1)}.$$

Let us consider $\mathbf{R}\boldsymbol{\theta} \geq \mathbf{0}_{(I-1)(J-1)}$, with $\mathbf{R} = \mathbf{e}_J \otimes (\mathbf{G}_{I-1} \otimes \mathbf{G}_{J-1}) = (\mathbf{0}_{(I-1)(J-1) \times (J-1)}, \mathbf{G}_{I-1} \otimes \mathbf{G}_{J-1})$, $\mathbf{0}_{a \times b}$ is the $a \times b$ matrix of zeros and \mathbf{G}_h is a $h \times h$ matrix with 1-s in the main diagonal and -1-s in the upper superdiagonal. Such restriction are equivalent to condition (11). Observe that the restrictions can be expressed also as $\mathbf{R}_{12}\boldsymbol{\theta}_{12} \geq \mathbf{0}_{(I-1)(J-1)}$, and $\boldsymbol{\theta}_2$ is a nuisance parameter vector because it does not take part actively in the restrictions. The kernel of the likelihood function with the new parametrization is obtained replacing \mathbf{p} by $\mathbf{p}(\boldsymbol{\theta})$ in (13), i.e.

$$\ell(\mathbf{N}; \boldsymbol{\theta}) = \mathbf{N}^T \log \mathbf{p}(\boldsymbol{\theta}) = \mathbf{N}^T \mathbf{W}_0 (\mathbf{u}(\tilde{\boldsymbol{\theta}}) - \mathbf{u}(\hat{\boldsymbol{\theta}})) + \mathbf{N}^T \mathbf{W} (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) = n u(\boldsymbol{\theta}) + \sum_{i=1}^{I-1} n_i u_{1(i)}(\boldsymbol{\theta}) + \mathbf{N}^T \mathbf{W} \boldsymbol{\theta}.$$

Hypotheses (2a)-(2b) or (3a)-(3b) can be now formulated as

$$H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{0}_{(I-1)(J-1)} \text{ versus } H_1 : \mathbf{R}\boldsymbol{\theta} \geq \mathbf{0}_{(I-1)(J-1)} \text{ and } \mathbf{R}\boldsymbol{\theta} \neq \mathbf{0}_{(I-1)(J-1)}. \quad (12)$$

3 Test-statistics based on phi-divergence measures

The likelihood function in our model is $\mathcal{L}(\mathbf{N}; \mathbf{p}) = \prod_{i=1}^I k_i \prod_{j=1}^J \pi_{ij}^{N_{ij}}$, where $k_i = n_i! / \prod_{j=1}^J N_{ij}!$, and the kernel of the loglikelihood function, in terms of \mathbf{p} , is

$$\ell(\mathbf{N}; \mathbf{p}) = \sum_{i=1}^I \sum_{j=1}^J N_{ij} \log p_{ij}. \quad (13)$$

Under H_0 , the parameter space is $\Theta_0 = \{\boldsymbol{\theta} \in \mathbb{R}^{I(J-1)} : \mathbf{R}\boldsymbol{\theta} = \mathbf{0}_{(I-1)(J-1)}\}$ and the maximum likelihood estimator of $\boldsymbol{\theta}$ in Θ_0 is $\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta_0} \ell(\mathbf{N}; \boldsymbol{\theta})$. Under either H_0 or H_1 , the overall parameter space is $\Theta_1 = \{\boldsymbol{\theta} \in \mathbb{R}^{I(J-1)} : \mathbf{R}\boldsymbol{\theta} \geq \mathbf{0}_{(I-1)(J-1)}\}$ and the maximum likelihood estimator of $\boldsymbol{\theta}$ in Θ_1 is $\tilde{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta_1} \ell(\mathbf{N}; \boldsymbol{\theta})$. The likelihood ratio test-statistic for testing (12) is

$$G^2 = 2(\ell(\mathbf{N}; \tilde{\boldsymbol{\theta}}) - \ell(\mathbf{N}; \hat{\boldsymbol{\theta}})) = 2n(u(\tilde{\boldsymbol{\theta}}) - u(\hat{\boldsymbol{\theta}})) + 2n \sum_{i=1}^{I-1} (u_{1(i)}(\tilde{\boldsymbol{\theta}}) - u_{1(i)}(\hat{\boldsymbol{\theta}})) + 2\mathbf{N}^T \mathbf{W} (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}), \quad (14)$$

and the chi-square test-statistic for testing (12) is

$$X^2 = n \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij}(\hat{\boldsymbol{\theta}}) - p_{ij}(\tilde{\boldsymbol{\theta}}))^2}{p_{ij}(\hat{\boldsymbol{\theta}})}. \quad (15)$$

Let $\bar{\mathbf{p}} = \mathbf{N}/n$ the vector of relative frequencies,

$$d_{Kull}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^I \sum_{j=1}^J p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

the Kullback-Leibler divergence measure between two IJ -dimensional probability vectors \mathbf{p} and \mathbf{q} , and

$$d_{Pearson}(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - q_{ij})^2}{q_{ij}}$$

the Pearson divergence measure. It is not difficult to check that $G^2 = 2n(d_{Kull}(\bar{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}})) - d_{Kull}(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}})))$ and $X^2 = 2nd_{Pearson}(\mathbf{p}(\hat{\boldsymbol{\theta}}), \mathbf{p}(\tilde{\boldsymbol{\theta}}))$. More general than the Kullback-Leibler divergence and Pearson divergence measures are ϕ -divergence measures, defined as

$$d_{\phi}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^I \sum_{j=1}^J q_{ij} \phi \left(\frac{p_{ij}}{q_{ij}} \right), \quad (16)$$

where $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ is a convex function such that $\phi(1) = \phi'(1) = 0$, $\phi''(1) > 0$, $0\phi(\frac{0}{0}) = 0$, $0\phi(\frac{p}{0}) = p \lim_{u \rightarrow \infty} \frac{\phi(u)}{u}$, for $p \neq 0$. For more details about ϕ -divergence measures see Pardo (2005).

Our second aim in this paper is to formulate test statistics valid for testing (12). Apart from the likelihood ratio statistic (14) and the chi-square (15) statistic, we shall consider two family of test-statistics based on ϕ -divergence measures,

$$T_{\phi}(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = \frac{2n}{\phi''(1)} (d_{\phi}(\bar{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}})) - d_{\phi}(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}))) \quad (17)$$

and

$$S_{\phi}(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = \frac{2n}{\phi''(1)} d_{\phi}(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})). \quad (18)$$

The two previous families of test-statistics can be considered as a natural extension of likelihood ratio test-statistic and chi-square test-statistic respectively. More thoroughly, for $\phi(x) = x \log x - x + 1$ in (17), we get the likelihood ratio test-statistic and for $\phi(x) = \frac{1}{2}(x - 1)^2$ in (18) we get the chi-square test-statistic.

Section 4 is devoted to present the main theoretical results in the paper. In Section 5, an example illustrates the results of Section 4. A simulation study is developed in Section 6, in order to study the behaviour of the families of test-statistics introduced in (17) and (18). Finally, we present an appendix in which we establish some parts of the proofs of the results given in Section 4.

4 Asymptotic distribution

For product-multinomial sampling, if we consider the partitioned matrix $\mathbf{W}^T = (\mathbf{W}_1^T, \dots, \mathbf{W}_I^T)$, such that $\log \mathbf{p}_i(\boldsymbol{\theta}) = u\mathbf{1}_J + u_{1(i)}\mathbf{1}_J + \mathbf{W}_i\boldsymbol{\theta}$, $i = 1, \dots, I-1$, $\log \mathbf{p}_I(\boldsymbol{\theta}) = u\mathbf{1}_J + \mathbf{W}_I\boldsymbol{\theta}$, and $\mathcal{I}_{F,i}^{(n_1, \dots, n_I)}(\boldsymbol{\theta}) = \mathbf{W}_i^T (\mathbf{D}_{\boldsymbol{\pi}_i(\boldsymbol{\theta})} - \boldsymbol{\pi}_i(\boldsymbol{\theta})\boldsymbol{\pi}_i^T(\boldsymbol{\theta}))\mathbf{W}_i$, we have

$$\mathcal{I}_F^{(n_1, \dots, n_I)}(\boldsymbol{\theta}) = \sum_{i=1}^I \frac{n_i}{n} \mathcal{I}_{F,i}^{(n_1, \dots, n_I)}(\boldsymbol{\theta}) = \mathbf{W}^T \left(\bigoplus_{i=1}^I \frac{n_i}{n} (\mathbf{D}_{\boldsymbol{\pi}_i(\boldsymbol{\theta})} - \boldsymbol{\pi}_i(\boldsymbol{\theta})\boldsymbol{\pi}_i^T(\boldsymbol{\theta})) \right) \mathbf{W}, \quad (19)$$

where $\bigoplus_{h=1}^a \mathbf{A}_h$ denotes the direct sum between the matrices $\{\mathbf{A}_h\}_{h=1}^a$. Our interest is to establish the Fisher information matrix, for $n \rightarrow \infty$, under the assumption

$$\boldsymbol{\pi}_1(\boldsymbol{\theta}_0) = \dots = \boldsymbol{\pi}_I(\boldsymbol{\theta}_0) = (\pi_1(\boldsymbol{\theta}_0), \dots, \pi_J(\boldsymbol{\theta}_0))^T = \boldsymbol{\pi}(\boldsymbol{\theta}_0), \quad i = 1, \dots, I,$$

which is equivalent to the null hypothesis (2a) or (3a). Let $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0) = (\pi_1(\boldsymbol{\theta}_0), \dots, \pi_{J-1}(\boldsymbol{\theta}_0))^T$ be the subvector of $\boldsymbol{\pi}(\boldsymbol{\theta}_0)$ obtained deleting the J -th element, $\pi_J(\boldsymbol{\theta}_0)$, from $\boldsymbol{\pi}(\boldsymbol{\theta}_0)$. If we consider the probability vector

$$\boldsymbol{\nu} = (\nu_1, \dots, \nu_I)^T,$$

such that $\nu_i = \lim_{n \rightarrow \infty} \frac{n_i}{n}$, $i = 1, \dots, I$, we denote by $\boldsymbol{\nu}^* = (\nu_1, \dots, \nu_{I-1})^T$ its subvector obtained deleting the I -th element, ν_I , from $\boldsymbol{\nu}$.

Theorem 1 *If we denote $\mathcal{I}_F(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \mathcal{I}_F^{(n_1, \dots, n_I)}(\boldsymbol{\theta})$ when $\boldsymbol{\theta} \in \Theta_0$, we have*

$$\mathcal{I}_F(\boldsymbol{\theta}_0) = \begin{pmatrix} 1 & \boldsymbol{\nu}^{*T} \\ \boldsymbol{\nu}^* & \bigoplus_{i=1}^{I-1} \nu_i \end{pmatrix} \otimes (\mathbf{D}_{\boldsymbol{\pi}^*(\boldsymbol{\theta}_0)} - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0)\boldsymbol{\pi}^{*T}(\boldsymbol{\theta}_0)). \quad (20)$$

Proof. Replacing $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_0$ and the explicit expression of \mathbf{W} in the general expression of the finite sample size Fisher information matrix for two independent multinomial samples, (19), we obtain through the property of the Kronecker product given in (1.22) of Harville (2008, page 341) that

$$\begin{aligned} \mathcal{I}_F^{(n_1, \dots, n_I)}(\boldsymbol{\theta}_0) &= \left(\begin{pmatrix} \mathbf{1}_{I-1} & \mathbf{I}_{I-1} \\ 1 & \mathbf{0}_{I-1}^T \end{pmatrix}^T \otimes \begin{pmatrix} \mathbf{I}_{J-1} \\ \mathbf{0}_{J-1}^T \end{pmatrix}^T \right) \left(\left(\bigoplus_{i=1}^I \frac{n_i}{n} \right) \otimes (\mathbf{D}_{\boldsymbol{\pi}(\boldsymbol{\theta}_0)} - \boldsymbol{\pi}(\boldsymbol{\theta}_0)\boldsymbol{\pi}^T(\boldsymbol{\theta}_0)) \right) \\ &\quad \times \left(\begin{pmatrix} \mathbf{1}_{I-1} & \mathbf{I}_{I-1} \\ 1 & \mathbf{0}_{I-1}^T \end{pmatrix} \otimes \begin{pmatrix} \mathbf{I}_{J-1} \\ \mathbf{0}_{J-1}^T \end{pmatrix} \right) \\ &= \left(\begin{pmatrix} \mathbf{1}_{I-1} & \mathbf{I}_{I-1} \\ 1 & \mathbf{0}_{I-1}^T \end{pmatrix}^T \begin{pmatrix} \bigoplus_{i=1}^I \frac{n_i}{n} \\ 1 \end{pmatrix} \begin{pmatrix} \mathbf{1}_{I-1} & \mathbf{I}_{I-1} \\ 1 & \mathbf{0}_{I-1}^T \end{pmatrix} \right) \otimes \left(\begin{pmatrix} \mathbf{I}_{J-1} \\ \mathbf{0}_{J-1}^T \end{pmatrix}^T (\mathbf{D}_{\boldsymbol{\pi}(\boldsymbol{\theta}_0)} - \boldsymbol{\pi}(\boldsymbol{\theta}_0)\boldsymbol{\pi}^T(\boldsymbol{\theta}_0)) \begin{pmatrix} \mathbf{I}_{J-1} \\ \mathbf{0}_{J-1}^T \end{pmatrix} \right) \\ &= \begin{pmatrix} 1 & (\frac{n_1}{n}, \dots, \frac{n_{I-1}}{n}) \\ (\frac{n_1}{n}, \dots, \frac{n_{I-1}}{n})^T & \left(\bigoplus_{i=1}^{I-1} \frac{n_i}{n} \right) \end{pmatrix} \otimes (\mathbf{D}_{\boldsymbol{\pi}^*(\boldsymbol{\theta}_0)} - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0)\boldsymbol{\pi}^{*T}(\boldsymbol{\theta}_0)), \end{aligned}$$

and then (20). ■

In the following theorem we present the asymptotic distribution of all of the proposed test-statistics under the null hypothesis. Let $E = \{1, \dots, (I-1)(J-1)\}$ the whole set of all row-indices of matrix \mathbf{R} , $\mathcal{F}(E)$ the family of all possible subsets of E , and $\mathbf{R}(S)$ is a submatrix of \mathbf{R} with row-indices belonging to $S \in \mathcal{F}(E)$.

Theorem 2 *Under H_0 , the asymptotic distribution of $S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ and $T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ is*

$$\lim_{n \rightarrow \infty} \Pr \left(S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \leq x \right) = \lim_{n \rightarrow \infty} \Pr \left(T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \leq x \right) = \sum_{h=0}^{(I-1)(J-1)} w_h(\boldsymbol{\theta}_0) \Pr \left(\chi_{(I-1)(J-1)-h}^2 \leq x \right)$$

where θ_0 is the true value of the unknown parameter, $\chi_0^2 \equiv 0$,

$$w_j(\theta_0) = \sum_{S \in \mathcal{F}(E), \text{card}(S)=h} \Pr(\mathbf{Z}_1(S) \geq \mathbf{0}_h) \Pr(\mathbf{Z}_2(S) \geq \mathbf{0}_{(I-1)(J-1)-h}), \quad (21)$$

$\mathbf{Z}_1(S) \sim \mathcal{N}(\mathbf{0}_{\text{card}(S)}, \Sigma_1(\theta_0, S))$, $\mathbf{Z}_2(S) \sim \mathcal{N}(\mathbf{0}_{(I-1)(J-1)-\text{card}(S)}, \Sigma_2(\theta_0, S))$, with

$$\Sigma_1(\theta_0, S) = \mathbf{H}^{-1}(S, S, \theta_0), \quad (22)$$

$$\Sigma_2(\theta_0, S) = \mathbf{H}(S^C, S^C, \theta_0) - \mathbf{H}(S^C, S, \theta_0) \mathbf{H}^{-1}(S, S, \theta_0) \mathbf{H}^T(S^C, S, \theta_0), \quad (23)$$

$S^C = E - S$, $\mathbf{H}(S_1, S_2, \theta_0)$ is the matrix obtained deleting from $\mathbf{H}(\theta_0)$ the row indices not contained in S_1 , the column indices not contained in S_2 ,

$$\mathbf{H}(\theta_0) = \mathbf{K}(\nu) \otimes \mathbf{K}(\pi(\theta_0)) \quad (24)$$

is the $(I-1)(J-1) \times (I-1)(J-1)$ matrix which depends on the symmetric tridiagonal matrices $\mathbf{K}(\nu)$ and $\mathbf{K}(\pi(\theta_0))$ defined as

$$\mathbf{K}(\mathbf{q}) = \mathbf{G}_{K-1} \mathbf{D}_{\mathbf{q}^*}^{-1} \mathbf{G}_{K-1}^T + \frac{1}{q_K} \mathbf{e}_{K-1} \mathbf{e}_{K-1}^T = \begin{pmatrix} \frac{q_1+q_2}{q_1 q_2} & -\frac{1}{q_2} & & & \\ -\frac{1}{q_2} & \frac{q_2+q_3}{q_2 q_3} & -\frac{1}{q_3} & & \\ & -\frac{1}{q_3} & \frac{q_3+q_4}{q_3 q_4} & \ddots & \\ & & \ddots & \ddots & -\frac{1}{q_{K-1}} \\ & & & -\frac{1}{q_{K-1}} & \frac{q_{K-1}+q_K}{q_{K-1} q_K} \end{pmatrix}, \quad (25)$$

where $\mathbf{q}^* = (q_1, \dots, q_{K-1})^T$, the subvector of a probability vector $\mathbf{q} = (q_1, \dots, q_K)^T$.

Proof. By following similar arguments of Martín and Balakrishnan we obtain $\mathbf{H}(S, S, \theta_0) = \mathbf{R}(S) \mathcal{I}_F^{-1}(\theta_0) \mathbf{R}^T(S)$ (see Section A.3 of the Appendix). We shall here obtain the expression of

$$\begin{aligned} \mathbf{H}(\theta_0) &= \mathbf{H}(E, E, \theta_0) = \mathbf{R}(E) \mathcal{I}_F^{-1}(\theta_0) \mathbf{R}^T(E) \\ &= (\mathbf{e}_J \otimes (\mathbf{G}_{I-1} \otimes \mathbf{G}_{J-1})) \mathcal{I}_F^{-1}(\theta_0) (\mathbf{e}_J^T \otimes (\mathbf{G}_{I-1} \otimes \mathbf{G}_{J-1})^T), \end{aligned}$$

where

$$\begin{aligned} \mathcal{I}_F^{-1}(\theta_0) &= \begin{pmatrix} 1 & \nu^{*T} \\ \nu^* & \bigoplus_{i=1}^{I-1} \nu_i \end{pmatrix}^{-1} \otimes (\mathbf{D}_{\pi^*(\theta_0)} - \pi^*(\theta_0) \pi^{*T}(\theta_0))^{-1} \\ &= \begin{pmatrix} \frac{1}{\nu_I} & -\frac{1}{\nu_I} \mathbf{1}_{I-1}^T \\ -\frac{1}{\nu_I} \mathbf{1}_{I-1} & \mathbf{D}_{\nu^*}^{-1} + \frac{1}{\nu_I} \mathbf{1}_{I-1} \mathbf{1}_{I-1}^T \end{pmatrix} \otimes \left(\mathbf{D}_{\pi^*(\theta_0)}^{-1} + \frac{1}{\pi_J(\theta_0)} \mathbf{1}_{J-1} \mathbf{1}_{J-1}^T \right), \end{aligned}$$

and then

$$\begin{aligned} \mathbf{H}(\theta_0) &= (\mathbf{e}_J \otimes (\mathbf{G}_{I-1} \otimes \mathbf{G}_{J-1})) \left(\begin{pmatrix} \frac{1}{\nu_I} & -\frac{1}{\nu_I} \mathbf{1}_{I-1}^T \\ -\frac{1}{\nu_I} \mathbf{1}_{I-1} & \mathbf{D}_{\nu^*}^{-1} + \frac{1}{\nu_I} \mathbf{1}_{I-1} \mathbf{1}_{I-1}^T \end{pmatrix} \otimes \left(\mathbf{D}_{\pi^*(\theta_0)}^{-1} + \frac{1}{\pi_J(\theta_0)} \mathbf{1}_{J-1} \mathbf{1}_{J-1}^T \right) \right) \\ &\quad \times (\mathbf{e}_J^T \otimes (\mathbf{G}_{I-1} \otimes \mathbf{G}_{J-1})^T) \\ &= (\mathbf{G}_{I-1} \otimes \mathbf{G}_{J-1}) \left(\left(\mathbf{D}_{\nu^*}^{-1} + \frac{1}{\nu_I} \mathbf{1}_{I-1} \mathbf{1}_{I-1}^T \right) \otimes \left(\mathbf{D}_{\pi^*(\theta_0)}^{-1} + \frac{1}{\pi_J(\theta_0)} \mathbf{1}_{J-1} \mathbf{1}_{J-1}^T \right) \right) (\mathbf{G}_{I-1}^T \otimes \mathbf{G}_{J-1}^T) \\ &= \left(\mathbf{G}_{I-1} \mathbf{D}_{\nu^*}^{-1} \mathbf{G}_{I-1}^T + \frac{1}{\nu_I} \mathbf{e}_{I-1} \mathbf{e}_{I-1}^T \right) \otimes \left(\mathbf{G}_{J-1} \mathbf{D}_{\pi^*(\theta_0)}^{-1} \mathbf{G}_{J-1}^T + \frac{1}{\pi_J(\theta_0)} \mathbf{e}_{J-1} \mathbf{e}_{J-1}^T \right), \end{aligned}$$

and thus it holds (24). ■

We must take into account that even though there is an equality in (12) which is effective only for θ_{12} , the rest of the components of θ are nuisance parameters, and hence we have a composite null hypothesis which requires estimation of θ , through $\hat{\theta}$.

The following result is very useful in order to calculate the weights of the chi-bar distribution by using simulation experiments.

Corollary 3 *Under H_0 , the weights $w_j(\theta_0)$ of the asymptotic distribution of $S_\phi(\mathbf{p}(\tilde{\theta}), \mathbf{p}(\hat{\theta}))$ and $T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\theta}), \mathbf{p}(\hat{\theta}))$, given in Theorem 2, can be expressed as*

$$\begin{aligned} w_{(I-1)(J-1)-j}(\theta_0) &= w_j(\theta_0; (I-1)(J-1), \mathbf{H}^{-1}(\theta_0), \mathbb{R}_+^{(I-1)(J-1)}) \\ &= \Pr \left(\arg \min_{\zeta \in \mathbb{R}_+^{(I-1)(J-1)}} (\mathbf{Z} - \zeta)^T \mathbf{H}(\theta_0) (\mathbf{Z} - \zeta) \in \mathbb{R}_+^{(I-1)(J-1)}(j) \right), \end{aligned} \quad (26)$$

with $\mathbf{H}(\theta_0)$ given by (24) and

$$\mathbf{H}^{-1}(\theta_0) = \mathbf{K}^{-1}(\nu) \otimes \mathbf{K}^{-1}(\pi(\theta_0)), \quad (27)$$

which depends on

$$\mathbf{K}^{-1}(\mathbf{q}) = \mathbf{T}_{K-1}^T (\mathbf{D}_{\mathbf{q}^*} - \mathbf{q}^* \mathbf{q}^{*T}) \mathbf{T}_{K-1}, \quad (28)$$

$\mathbf{T}_h = \mathbf{G}_h^{-1}$ is an upper triangular matrix of 1-s, $\mathbf{Z} \sim \mathcal{N}_{(I-1)(J-1)}(\mathbf{0}_{(I-1)(J-1)}, \mathbf{H}^{-1}(\theta_0))$, and $\mathbb{R}_+^{(I-1)(J-1)}(j)$ is the subset of $\mathbb{R}_+^{(I-1)(J-1)} = \{\zeta \in \mathbb{R}^{(I-1)(J-1)} : \zeta \geq \mathbf{0}_{(I-1)(J-1)}\}$, such that j components of the $(I-1)(J-1)$ -dimensional vectors are strictly positive and $(I-1)(J-1) - j$ components are null.

Proof. It is well known that the weights of a chi-bar distribution can be interpreted in terms of the projection of a p -dimensional central normal distribution \mathbf{Z}_p with a non-singular variance-covariance matrix \mathbf{V} , on a closed convex cone in \mathbb{R}^p , C , as $w_j(p, \mathbf{V}, C) = \Pr(\Pi(\mathbf{Z}_p|C) \in \mathbb{R}_+^p(j))$ where

$$\Pi(\mathbf{Z}_p|C) = \arg \min_{\zeta \in C} (\mathbf{Z}_p - \zeta)^T \mathbf{V}^{-1} (\mathbf{Z}_p - \zeta).$$

Now, focussed on $w_j(\theta_0) = w_j(\theta_0; p, \mathbf{V}, C)$ in Theorem 2, we must identify the value of p , expression of matrix \mathbf{V} and the set C . In Kudô (1963, p.414) and also in Shapiro (1988, p.54) it is shown that

$$w_j(p, \mathbf{V}, \mathbb{R}_+^p) = \sum_{S \in \mathcal{F}(F), \text{card}(S)=p-j} \Pr(\mathbf{Z}_{1,p}(S) \geq \mathbf{0}_j) \Pr(\mathbf{Z}_{2,p}(S) \geq \mathbf{0}_{(I-1)(J-1)-j}),$$

where $F = \{1, \dots, p\}$, $\mathbf{Z}_{1,p}(S) \sim \mathcal{N}_{\text{card}(S)}(\mathbf{0}_{\text{card}(S)}, \mathbf{V}^{-1}(S))$ and $\mathbf{Z}_{2,p}(S) \sim \mathcal{N}_{\text{card}(F)-\text{card}(S)}(\mathbf{0}_{\text{card}(F)-\text{card}(S)}, \mathbf{V}(F-S, S))$, where $\mathbf{V}(S)$ is the variance-covariance matrix of the random vector obtained by considering only from \mathbf{Z}_p the components belonging to S and $\mathbf{V}(S, F-S)$ is the same but rather than ignoring the components out from S they are considered equals zero. This enunciate can be also seen in Silvapulle and Sen (2005, page 83). Note that we can identify $p = (I-1)(J-1)$, $\mathbf{V} = \mathbf{H}(\theta_0)$ and $C = \mathbb{R}_+^{(I-1)(J-1)}$ and

$$\begin{aligned} &w_{(I-1)(J-1)-j}(\theta_0; (I-1)(J-1), \mathbf{H}(\theta_0), \mathbb{R}_+^{(I-1)(J-1)}) \\ &= \sum_{S \in \mathcal{F}(F), \text{card}(S)=j} \Pr(\mathbf{Z}_{1,J-1}(S) \geq \mathbf{0}_j) \Pr(\mathbf{Z}_{2,J-1}(S) \geq \mathbf{0}_{(I-1)(J-1)-j}), \end{aligned}$$

which is equal to

$$\begin{aligned} &w_j(\theta_0; (I-1)(J-1), \mathbf{H}^{-1}(\theta_0), \mathbb{R}_+^{(I-1)(J-1)}) \\ &= \sum_{S \in \mathcal{F}(F), \text{card}(S)=j} \Pr(\mathbf{Z}_{1,J-1}(S) \geq \mathbf{0}_j) \Pr(\mathbf{Z}_{2,J-1}(S) \geq \mathbf{0}_{(I-1)(J-1)-j}), \end{aligned}$$

according to Proposition 3.6.1(7) in Silvapulle and Sen (2005, page 82). This expression match (21). ■

Since θ_0 is unknown, we cannot use directly the results based on Theorem 2 or Corollary 3. However, the unknown parameter θ_0 can be replaced by its estimator under the null hypothesis, $\hat{\theta}$. The tests performed replacing θ_0 by $\hat{\theta}$ are called “local tests” (see Dardanoni and Forcina (1998)) and they are usually considered to be good approximations of the theoretical tests. It is worthwhile to mention that $p(\hat{\theta})$ has an explicit expression,

$$p_{ij}(\hat{\theta}) = \hat{\nu}_i \pi_j(\hat{\theta}), \quad \hat{\nu}_i = \frac{n_i}{n}, \quad \pi_j(\hat{\theta}) = \frac{1}{n} N_{\bullet j}, \quad N_{\bullet j} = \sum_{h=1}^I N_{hj}. \quad (29)$$

Now, based on Corollary 3, and taking into account that (26) is equal to

$$w_j(\theta_0) = \Pr \left(\arg \min_{\zeta \in \mathbb{R}_+^{(I-1)(J-1)}} \frac{1}{2} \zeta^T \mathbf{H}(\theta_0) \zeta - (\mathbf{H}(\theta_0) \mathbf{Z})^T \zeta \in \mathbb{R}_+^{(I-1)(J-1)}(j) \right), \quad (30)$$

where $\mathbf{H}(\theta_0)$ is (24) and $\mathbf{Z} \sim \mathcal{N}_{(I-1)(J-1)}(\mathbf{0}_{(I-1)(J-1)}, \mathbf{H}^{-1}(\theta_0))$, we shall consider an algorithm for obtaining the weights associated with a sample.

Algorithm 4 (Estimation of weights) *The weights of the local tests, $w_j(\hat{\theta})$, are obtained by Monte Carlo, once we have a realization \mathbf{n} of \mathbf{N} in the following way*

STEP 1: Using \mathbf{n} , calculate ν and $\pi(\hat{\theta})$ taking into account (29).

STEP 2: Compute $\mathbf{H}(\hat{\theta})$ by following (24), in terms of $\mathbf{K}(\hat{\nu})$, $\mathbf{K}(\pi(\hat{\theta}))$, given by (25)).

STEP 3: Compute $\mathbf{H}^{-1}(\hat{\theta})$ by following (28), in terms of $\mathbf{K}^{-1}(\hat{\nu})$, $\mathbf{K}^{-1}(\pi(\hat{\theta}))$, given by (28).

STEP 4: For $j = 0, \dots, (I-1)(J-1)$, do $N(j) := 0$.

STEP 5: Repeat the following steps R (say $R = 1,000,000$) times:

STEP 5.1: Generate an observation, \mathbf{z} , from $\mathbf{Z} \sim \mathcal{N}_{(I-1)(J-1)}(\mathbf{0}_{(I-1)(J-1)}, \mathbf{H}^{-1}(\theta_0))$. E.g., the NAG Fortran library subroutines G05CBF, G05EAF, and G05EZF can be useful.

STEP 5.2: Compute $\hat{\zeta}(\mathbf{z}) = \arg \min_{\zeta \in \mathbb{R}_+^{(I-1)(J-1)}} \frac{1}{2} \zeta^T \mathbf{H}(\hat{\theta}) \zeta - (\mathbf{H}(\hat{\theta}) \mathbf{z})^T \zeta$. E.g., the IMSL Fortran library subroutine DQPROG can be useful.

STEP 5.3: Count j^ , the number of strictly positive components contained in $\hat{\zeta}(\mathbf{z})$, and do $N(j^*) := N(j^*) + 1$.*

STEP 6: Do $w_j(\hat{\theta}) := \frac{N(j)}{R}$ for $j = 0, \dots, (I-1)(J-1)$.

See <http://www.nag.co.uk/numeric/fl/FLdescription.asp>, for details about subroutines of the NAG Fortran library, and <http://www.roguewave.com/Portals/0/products/ims-l-numerical-libraries/fortran-library/docs/7.0/math/math.htm> for the IMSL Fortran library. It is worthwhile to mention that these values can be also computed using mvtnorm R package (see <http://CRAN.R-project.org/package=mvtnorm>, for details), however this method based on numerical integration tends to provide less accurate values.

5 Example

In this section we are going to analyze the data set of Section 1, using the proposed test-statistics. By following the specific notation of our paper, we are considering two ordinal variables associated with $n = 417$ duodenal ulcer patients in a hospital, X =severity of the operation, classified in an increasing order from 1 to $I = 4$, and Y =extent of side effects, categorized as None (1), Slight (2) and Moderate ($J = 3$). The sample, a realization of \mathbf{N} , is summarized in

$$\begin{aligned} \mathbf{n} &= (n_{11}, n_{12}, n_{13}, n_{21}, n_{22}, n_{23}, n_{31}, n_{32}, n_{33}, n_{41}, n_{42}, n_{43})^T \\ &= (61, 28, 7, 68, 23, 13, 58, 40, 12, 53, 38, 16)^T. \end{aligned}$$

The order restricted maximum likelihood estimator (MLE) of $\boldsymbol{\theta} = (\boldsymbol{\theta}_2^T, \boldsymbol{\theta}_{12}^T)^T$ under likelihood ratio order, obtained through E04UCF subroutine of NAG Fortran library (<http://www.nag.co.uk/numeric/fl/FLdescription.asp>), is $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\theta}}_2^T, \tilde{\boldsymbol{\theta}}_{12}^T)^T$, with

$$\tilde{\boldsymbol{\theta}}_2 = (1.1977, 0.8650)^T, \quad \tilde{\boldsymbol{\theta}}_{12} = (0.9983, 0.4501, 0.6376, 0.0894, 0.1916, 0.0894)^T.$$

The estimated probability vectors of interest are

$$\hat{\boldsymbol{\nu}} = \left(\frac{n_1}{n}, \frac{n_2}{n}, \frac{n_3}{n}, \frac{n_4}{n} \right)^T = \left(\frac{96}{417}, \frac{104}{417}, \frac{110}{417}, \frac{107}{417} \right)^T,$$

$$\boldsymbol{\pi}(\hat{\boldsymbol{\theta}}) = \left(\frac{n_{\bullet 1}}{n}, \frac{n_{\bullet 2}}{n}, \frac{n_{\bullet 3}}{n} \right)^T = \left(\frac{240}{417}, \frac{129}{417}, \frac{48}{417} \right)^T,$$

$$\bar{\mathbf{p}} = (0.1463, 0.0671, 0.0168, 0.1631, 0.0552, 0.0312, 0.1391, 0.0959, 0.0288, 0.1271, 0.0911, 0.0384)^T,$$

$$\mathbf{p}(\tilde{\boldsymbol{\theta}}) = (0.1509, 0.0625, 0.0168, 0.1585, 0.0657, 0.0253, 0.1391, 0.0900, 0.0347, 0.1271, 0.0911, 0.0384)^T,$$

$$\mathbf{p}(\hat{\boldsymbol{\theta}}) = (0.1325, 0.0712, 0.0265, 0.1435, 0.0772, 0.0287, 0.1518, 0.0816, 0.0304, 0.1477, 0.0794, 0.0295)^T,$$

and the weights

$$\{w_j(\hat{\boldsymbol{\theta}})\}_{j=0}^6 = \{0.0006103103, 0.009753533, 0.06122672, 0.1953851, 0.3353725, 0.2949007, 0.1028136\}, \quad (31)$$

were obtained using Theorem 2 and the R package mvtnorm (located at <http://cran.r-project.org/web/packages/mvtnorm/index.html>), once we knew

$$\mathbf{K}(\hat{\boldsymbol{\nu}}) = \begin{pmatrix} n \frac{n_1+n_2}{n_1 n_2} & -\frac{n}{n_2} & 0 \\ -\frac{n}{n_2} & n \frac{n_2+n_3}{n_2 n_3} & -\frac{n}{n_3} \\ 0 & -\frac{n}{n_3} & n \frac{n_3+n_4}{n_3 n_4} \end{pmatrix} = \begin{pmatrix} \frac{3475}{416} & -\frac{417}{417} & 0 \\ -\frac{417}{417} & \frac{44619}{5720} & -\frac{417}{110} \\ 0 & -\frac{417}{110} & \frac{90489}{11770} \end{pmatrix},$$

$$\mathbf{K}(\boldsymbol{\pi}(\hat{\boldsymbol{\theta}})) = \begin{pmatrix} n \frac{n_{\bullet 1}+n_{\bullet 2}}{n_{\bullet 1} n_{\bullet 2}} & -\frac{n}{n_{\bullet 2}} \\ -\frac{n}{n_{\bullet 2}} & n \frac{n_{\bullet 2}+n_{\bullet 3}}{n_{\bullet 2} n_{\bullet 3}} \end{pmatrix} = \begin{pmatrix} \frac{17097}{3440} & -\frac{417}{129} \\ -\frac{417}{129} & \frac{8201}{688} \end{pmatrix},$$

$$\mathbf{H}(\hat{\boldsymbol{\theta}}) = \mathbf{K}(\hat{\boldsymbol{\nu}}) \otimes \mathbf{K}(\boldsymbol{\pi}(\hat{\boldsymbol{\theta}})) = \begin{pmatrix} \frac{11882415}{286208} & -\frac{483025}{28498475} & -\frac{7129449}{357760} & \frac{57963}{4472} & 0 & 0 \\ -\frac{483025}{28498475} & \frac{286208}{57963} & -\frac{357760}{3419817} & -\frac{4472}{3419817} & 0 & 0 \\ -\frac{7129449}{357760} & -\frac{357760}{3419817} & \frac{762851043}{6202041} & -\frac{71552}{6202041} & -\frac{7129449}{378400} & \frac{57963}{4730} \\ \frac{57963}{4472} & -\frac{3419817}{71552} & -\frac{6202041}{245960} & \frac{365920419}{3935360} & -\frac{378400}{4730} & -\frac{4730}{75680} \\ 0 & 0 & -\frac{7129449}{378400} & \frac{3935360}{57963} & \frac{1547090433}{506110} & -\frac{12577971}{742100289} \\ 0 & 0 & \frac{378400}{4730} & -\frac{3419817}{75680} & -\frac{40488800}{12577971} & \frac{506110}{8097760} \end{pmatrix},$$

Along the current section, we are trying to express the matrices as precise as possible in order to highlight that the proposed method provide very simple accurate way for obtaining the weights even for big dimensions. In the notation we understand that $n_{\bullet j}$, $j = 1, 2, 3$, are realizations of the r.v. $N_{\bullet j}$ defined in (29). The output of the code based on such package, provides normal orthant probabilities based on numerical integration, as well as the precision error. Taking into account Proposition 3.6.1(3) in Silvapulle and Sen (2005, page 82), $\sum_{i=0}^6 (-1)^i w_i(\hat{\boldsymbol{\theta}}) = 0$ should be held theoretically, and for (31) we obtained $\sum_{i=0}^6 (-1)^i w_i(\hat{\boldsymbol{\theta}}) = -1.6203 \times 10^{-5}$, and this means that 1.6203×10^{-5} could be considered as an overall measure of precision error for the weights.

Using Algorithm 4 and taking into account

$$\begin{aligned}
\mathbf{K}^{-1}(\hat{\nu}) &= \mathbf{T}_3^T \left(\mathbf{D}_{\hat{\nu}^*} - \hat{\nu}^* \hat{\nu}^{*T} \right) \mathbf{T}_3 = \begin{pmatrix} \frac{3424}{19321} & \frac{6944}{57963} & \frac{3424}{57963} \\ \frac{6944}{57963} & \frac{173889}{173889} & \frac{173889}{173889} \\ \frac{3424}{57963} & \frac{173889}{173889} & \frac{173889}{173889} \end{pmatrix} \\
\mathbf{K}^{-1}(\pi(\hat{\theta})) &= \mathbf{T}_2^T \left(\mathbf{D}_{\pi^*(\hat{\theta})} - \pi^*(\hat{\theta}) \pi^{*T}(\hat{\theta}) \right) \mathbf{T}_2 = \begin{pmatrix} \frac{4720}{19321} & \frac{1280}{19321} \\ \frac{1280}{19321} & \frac{1280}{19321} \end{pmatrix} \\
\mathbf{H}^{-1}(\hat{\theta}) &= \mathbf{K}^{-1}(\hat{\nu}^*) \otimes \mathbf{K}^{-1}(\pi(\hat{\theta})) = \\
&= \begin{pmatrix} \frac{16161280}{4382720} & \frac{4382720}{373301041} & \frac{32775680}{1119903123} & \frac{8888320}{1119903123} & \frac{16161280}{1119903123} & \frac{4382720}{1119903123} \\ \frac{4382720}{373301041} & \frac{373301041}{6738432} & \frac{1119903123}{8888320} & \frac{1119903123}{4555264} & \frac{1119903123}{4382720} & \frac{1119903123}{2246144} \\ \frac{32775680}{1119903123} & \frac{6738432}{8888320} & \frac{1119903123}{204848000} & \frac{373301041}{55552000} & \frac{1119903123}{101008000} & \frac{373301041}{27392000} \\ \frac{8888320}{1119903123} & \frac{1119903123}{4555264} & \frac{3359709369}{55552000} & \frac{3359709369}{28470400} & \frac{3359709369}{27392000} & \frac{3359709369}{14038400} \\ \frac{16161280}{1119903123} & \frac{4382720}{373301041} & \frac{3359709369}{101008000} & \frac{1119903123}{27392000} & \frac{3359709369}{156562400} & \frac{1119903123}{42457600} \\ \frac{4382720}{1119903123} & \frac{1119903123}{2246144} & \frac{3359709369}{27392000} & \frac{3359709369}{14038400} & \frac{3359709369}{42457600} & \frac{3359709369}{21759320} \end{pmatrix},
\end{aligned}$$

very similar weights were obtained: $\{w_j(\hat{\theta})\}_{j=0}^6 = \{0.000613, 0.009627, 0.060873, 0.195312, 0.335389, 0.295527, 0.102659\}$. From these weights the quantile of order 0.05, which defines the rejection region, was find to be 6.34.

If we take, for (16), $\phi_\lambda(x) = \frac{1}{\lambda(1+\lambda)}(x^{\lambda+1} - x - \lambda(x-1))$, where for each $\lambda \in \mathbb{R} - \{-1, 0\}$ a different divergence measure is constructed, a very important subfamily called “power divergence family of measures” is obtained

$$d_\lambda(\mathbf{p}, \mathbf{q}) = \frac{1}{\lambda(\lambda+1)} \left(\sum_{i=1}^I \sum_{j=1}^J \frac{p_{ij}^{\lambda+1}}{q_{ij}^\lambda} - 1 \right), \text{ for each } \lambda \in \mathbb{R} - \{-1, 0\}. \quad (32)$$

It is also possible to cover the real line for λ , by defining $d_\lambda(\mathbf{p}, \mathbf{q}) = \lim_{t \rightarrow \lambda} d_t(\mathbf{p}, \mathbf{q})$, for $\lambda \in \{-1, 0\}$. It is well known that $d_0(\mathbf{p}, \mathbf{q}) = d_{Kull}(\mathbf{p}, \mathbf{q})$ and $d_1(\mathbf{p}, \mathbf{q}) = d_{Pearson}(\mathbf{p}, \mathbf{q})$, which is very interesting because the power divergence based family of test-statistics, which contains as special cases G^2 and X^2 , can be created. It is also worthwhile to mention that $d_{-1}(\mathbf{p}, \mathbf{q}) = d_{Kull}(\mathbf{q}, \mathbf{p})$.

When the test-statistic (17) and (18), based on power-divergences (32), are applied we get

$$T_\lambda(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\theta}), \mathbf{p}(\hat{\theta})) = 2n(d_\lambda(\bar{\mathbf{p}}, \mathbf{p}(\hat{\theta})) - d_\lambda(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\theta}))) = \frac{2n}{\lambda(\lambda+1)} \left(\sum_{i=1}^I \sum_{j=1}^J \frac{\bar{p}_{ij}^{\lambda+1}}{p_{ij}^\lambda(\hat{\theta})} - \sum_{i=1}^I \sum_{j=1}^J \frac{\bar{p}_{ij}^{\lambda+1}}{p_{ij}^\lambda(\tilde{\theta})} \right) \quad (33)$$

and

$$S_\lambda(\mathbf{p}(\tilde{\theta}), \mathbf{p}(\hat{\theta})) = 2nd_\lambda(\mathbf{p}(\tilde{\theta}), \mathbf{p}(\hat{\theta})) = \frac{2n}{\lambda(\lambda+1)} \left(\sum_{i=1}^I \sum_{j=1}^J \frac{p_{ij}^{\lambda+1}(\tilde{\theta})}{p_{ij}^\lambda(\hat{\theta})} - 1 \right), \quad (34)$$

for $\lambda \in \mathbb{R} - \{0, -1\}$, and $T_\lambda(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\theta}), \mathbf{p}(\hat{\theta})) = \lim_{\lambda \rightarrow \ell} T_\ell(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\theta}), \mathbf{p}(\hat{\theta}))$, $S_\lambda(\mathbf{p}(\tilde{\theta}), \mathbf{p}(\hat{\theta})) = \lim_{\lambda \rightarrow \ell} S_\ell(\mathbf{p}(\tilde{\theta}), \mathbf{p}(\hat{\theta}))$, $\lambda \in \{0, -1\}$, i.e.

$$T_0(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\theta}), \mathbf{p}(\hat{\theta})) = 2n(d_{Kull}(\bar{\mathbf{p}}, \mathbf{p}(\hat{\theta})) - d_{Kull}(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\theta}))) = 2n \sum_{i=1}^I \sum_{j=1}^J \bar{p}_{ij} \log \left(\frac{p_{ij}(\tilde{\theta})}{p_{ij}(\hat{\theta})} \right), \quad (35)$$

$$\begin{aligned}
T_{-1}(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\theta}), \mathbf{p}(\hat{\theta})) &= 2n(d_{Kull}(\mathbf{p}(\hat{\theta}), \bar{\mathbf{p}}) - d_{Kull}(\mathbf{p}(\tilde{\theta}), \bar{\mathbf{p}})) \\
&= 2n \left(\sum_{i=1}^I \sum_{j=1}^J p_{ij}(\hat{\theta}) \log \left(\frac{p_{ij}(\hat{\theta})}{\bar{p}_{ij}} \right) - \sum_{i=1}^I \sum_{j=1}^J p_{ij}(\tilde{\theta}) \log \left(\frac{p_{ij}(\tilde{\theta})}{\bar{p}_{ij}} \right) \right) \quad (36)
\end{aligned}$$

and

$$S_0(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = 2nd_{Kull}(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = 2n \sum_{i=1}^I \sum_{j=1}^J p_{ij}(\tilde{\boldsymbol{\theta}}) \log \left(\frac{p_{ij}(\tilde{\boldsymbol{\theta}})}{p_{ij}(\hat{\boldsymbol{\theta}})} \right), \quad (37)$$

$$S_{-1}(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = 2nd_{Kull}(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = 2n \sum_{i=1}^I \sum_{j=1}^J p_{ij}(\hat{\boldsymbol{\theta}}) \log \left(\frac{p_{ij}(\tilde{\boldsymbol{\theta}})}{p_{ij}(\hat{\boldsymbol{\theta}})} \right). \quad (38)$$

Suppose we want to consider a set of values for the parameter λ , Λ . The power divergence based test-statistics cover as special cases the classical ones (14), (15), actually $T_0(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = G^2$ and $S_1(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = X^2$. The power divergence based test-statistics with $\lambda = \frac{2}{3}$ are commonly considered for analysis because their performance is usually quite good. At this setting, a possible choice for studying its p -values is $\lambda \in \Lambda = \{-1.5, -1, -0.5, 0, \frac{2}{3}, 1, 1.5, 2\}$. We shall consider an algorithm for obtaining the p -values associated with hypothesis testing (12) for a given sample.

Algorithm 5 (Calculation of p -value) Let $T \in \{T_\lambda(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})), S_\lambda(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))\}_{\lambda \in \Lambda}$ be the test-statistic associated with (12). In the following steps the corresponding asymptotic p -value is calculated once it is suppose we have $\{w_j(\hat{\boldsymbol{\theta}})\}_{j=0}^{(I-1)(J-1)}$:

STEP 1: Using n calculate $\mathbf{p}(\hat{\boldsymbol{\theta}})$ taking into account (29).

STEP 2: Using $\mathbf{p}(\hat{\boldsymbol{\theta}})$ calculate value t of test-statistic T using the corresponding expression in (33)–(38).

STEP 3: Compute $p\text{-value}(T) := 0$.

STEP 4: If $t \leq 0$, do $p\text{-value}(T) := 1$, otherwise (if $t > 0$)

for $h = 0, \dots, (I-1)(J-1) - 1$, do $p\text{-value}(T) := p\text{-value}(T) + w_h(\hat{\boldsymbol{\theta}}) \Pr \left(\chi_{(I-1)(J-1)-h}^2 > t \right)$.

E.g., the NAG Fortran library subroutine G01ECF can be useful.

(Remark: for small sample sizes and for values of $T_\lambda(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$, sometimes $t < 0$).

In Table 2, the power divergence based test-statistics and their corresponding asymptotic p -values are shown. For all the power divergence based test-statistics, $T \in \{T_\lambda(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})), S_\lambda(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))\}_{\lambda \in \Lambda}$, the order restricted hypothesis cannot be rejected with a significance level 0.05. So, it is accepted that the probability of having side effects increases when the severity degree of the operation increases.

test-statistic	$\lambda = -1.5$	$\lambda = -1$	$\lambda = -0.5$	$\lambda = 0$	$\lambda = \frac{2}{3}$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$
$T_\lambda(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$	9.4681	9.1918	8.9535	8.7497	8.5262	8.4334	8.3160	8.2230
$p\text{-value}(T_\lambda)$	0.0123	0.0139	0.0155	0.0170	0.0188	0.0196	0.0206	0.0215
$S_\lambda(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$	9.1282	8.9774	8.8520	8.7497	8.6463	8.6076	8.5650	8.5399
$p\text{-value}(S_\lambda)$	0.0143	0.0153	0.0162	0.0170	0.0178	0.0181	0.0184	0.0186

Table 2: Power divergence based test-statistics and asymptotic p -values.

6 Monte Carlo Study

Taking five cases, depending on $\delta \in \{0, 0.1, 0.5, 1, 1.5, 15\}$, we considered $I = 4$ independent trinomial samples ($J = 3$) with a vector of theoretical probabilities,

$$\begin{aligned} \boldsymbol{\pi}_i(\boldsymbol{\theta}(\delta)) &= (\pi_{i1}(\boldsymbol{\theta}(\delta)), \pi_{i2}(\boldsymbol{\theta}(\delta)), \pi_{i3}(\boldsymbol{\theta}(\delta)))^T \\ \pi_{ij}(\boldsymbol{\theta}(\delta)) &= \frac{1}{3} \frac{1 + i(j-1)\delta}{1 + i\delta}, \quad i = 1, \dots, 4, \quad j = 1, \dots, 3, \end{aligned}$$

for each of the $I = 4$ independent multinomial samples, in four scenarios:

- * Scenario 1: $n = 28, n_1 = 4, n_2 = 6, n_3 = 8, n_4 = 10$;
- * Scenario 2: $n = 56, n_1 = 8, n_2 = 12, n_3 = 16, n_4 = 20$;
- * Scenario 3: $n = 84, n_1 = 12, n_2 = 18, n_3 = 24, n_4 = 30$;
- * Scenario 4: $n = 112, n_1 = 16, n_2 = 24, n_3 = 32, n_4 = 40$.

It is worthwhile to mention that we have chosen either equal or unequal sample sizes and we did not find any different performance as it was found for the stochastic ordering in Wang (1996). When $\delta = 0$, the null hypothesis is held, $\pi_i(\boldsymbol{\theta}(0)) = \boldsymbol{\pi}(\boldsymbol{\theta}_0) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T, i = 1, 2, 3, 4$, while in the rest of the values of δ the alternative hypothesis is held.

Let $R = 10,000$ be the number of replications considered for the Monte Carlo study. Once the nominal size of the test is prefixed to be $\alpha = 0.05$, the exact size of the test associated with $T \in \{T_\lambda(\bar{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}), \mathbf{p}(\hat{\boldsymbol{\theta}})), S_\lambda(\mathbf{p}(\boldsymbol{\theta}), \mathbf{p}(\hat{\boldsymbol{\theta}}))\}_{\lambda \in \Lambda}$, $\lambda \in \Lambda = \{-1.5, -1, -0.5, 0, \frac{2}{3}, 1, 1.5, 2\}$, can be estimated through

$$\hat{\alpha}_T = \frac{\sum_{h=1}^R I(p\text{-value}(T_h) < \alpha)}{R},$$

taking into account that $p\text{-value}(T_h)$ is the p -value obtained in the h -th replication by using Algorithm 5 and $I(\bullet)$ is the indicator function, which takes value 1 if \bullet is true and 0 otherwise. It is expected a more precise value of $\hat{\alpha}_T$ with respect to the nominal size α , as n is greater (the least precise nominal sizes in Scenario 1 and the most precise nominal sizes in Scenario 4). The first interest of the simulation study is focussed on identifying which test-statistic has the best approximation of $\hat{\alpha}_T$ with respect to α , for all the scenarios.

In Table 3 the local odds ratios,

$$\vartheta_{ij} = \vartheta_{ij}(\delta) = \frac{1 + i(j-1)\delta}{1 + (i+1)(j-1)\delta} \frac{1 + (i+1)j\delta}{1 + ij\delta},$$

$(i, j) \in \{1, 2, 3\} \times \{1, 2\}$, are shown for $\delta \in \{0.1, 0.5, 1, 1.5\}$. Notice that in $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}(\delta) = (\vartheta_{11}(\delta), \vartheta_{12}(\delta), \vartheta_{21}(\delta), \vartheta_{22}(\delta), \vartheta_{31}(\delta), \vartheta_{32}(\delta))^T$ some of the components are further from $\boldsymbol{\vartheta}(0) = \mathbf{1}_6$ (null hypothesis), as the value of $\delta > 0$ is further from 0. This means that a greater value of the estimation of the power function might be obtained,

$$\hat{\beta}_T(\delta) = \frac{\sum_{h=1}^R I(p\text{-value}(T_h) < \alpha)}{R},$$

as $\delta > 0$ is greater. This claim is supported by the fact that some values of the components of $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}(\delta)$ decrease as δ increases but more slowly than the others increase. In addition, for a fixed value of $\delta > 0$, it is expected a greater value of $\hat{\beta}_T(\delta)$, as n is greater (the worst powers in Scenario 1 and the best powers in Scenario 4). It is also worthwhile to mention that as δ increases $\pi_{i+1,1}(\boldsymbol{\theta}(\delta))/\pi_{i1}(\boldsymbol{\theta}(\delta))$ remains constant, $\pi_{i+1,2}(\boldsymbol{\theta}(\delta))/\pi_{i2}(\boldsymbol{\theta}(\delta))$ is not constant for $i = 1, 2, 3$, and $\pi_{i+1,2}(\boldsymbol{\theta}(\delta))/\pi_{i2}(\boldsymbol{\theta}(\delta))$ is approaching the limit $(\pi_{i+1,2}(\boldsymbol{\theta}(\infty))/\pi_{i2}(\boldsymbol{\theta}(\infty)))$ on the right for $i = 1, 2, 3$. The second interest of the simulation study is focussed in identifying which test-statistic has the best performance in powers and at the same time in approximating $\hat{\alpha}_T$ by α , in all the scenarios.

Once a nominal size $\alpha = 0.05$ is established, Table 4 summarizes the simulated exact sizes in all the scenarios for the test-statistic $T \in \{T_\lambda, S_\lambda\}_{\lambda \in \Lambda}$, with $\Lambda = \{-1.5, -1, -\frac{1}{2}, 0, \frac{2}{3}, 1, 1.5, 2\}$. We have plotted 3×2 graphs in Figures 1-4 and we refer them as plots in three rows. In the first row of Figures 1-4 we can see on the left the exact power in all the scenarios for the test-statistic $\{T_\lambda\}_{\lambda \in [-1.5, 3]}$ and on the right for the test-statistic $\{S_\lambda\}_{\lambda \in [-1.5, 3]}$. In order to make a comparison of exact powers, we cannot directly proceed without considering the exact sizes. For this reason we are going to give a procedure based on two steps.

Step 1: We are going to check for all the power divergence based test-statistics the criterion given by Dale (1986), i.e.,

$$|\text{logit}(1 - \hat{\alpha}_T) - \text{logit}(1 - \alpha)| \leq e \quad (39)$$

with $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$. We only consider the values of λ such that $\hat{\alpha}_T$ verifies (39) with $e = 0.35$, then we shall only consider the test-statistics such that $\hat{\alpha}_T \in [0.0357, 0.0695]$, in all the scenarios. This criterion has

	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = \infty$
$\vartheta_{11} = \vartheta_{11}(\delta)$	1.000	1.091	1.333	1.500	1.600	2.00
$\vartheta_{12} = \vartheta_{12}(\delta)$	1.000	1.069	1.125	1.111	1.094	1.00
$\vartheta_{21} = \vartheta_{21}(\delta)$	1.000	1.083	1.250	1.333	1.375	1.50
$\vartheta_{22} = \vartheta_{22}(\delta)$	1.000	1.055	1.066	1.050	1.039	1.00
$\vartheta_{31} = \vartheta_{31}(\delta)$	1.000	1.077	1.200	1.250	1.273	1.33
$\vartheta_{32} = \vartheta_{32}(\delta)$	1.000	1.045	1.042	1.029	1.021	1.000
$\pi_{21}(\boldsymbol{\theta}(\delta))/\pi_{11}(\boldsymbol{\theta}(\delta))$	0.33/0.33	0.28/0.30	0.17/0.22	0.11/0.17	0.08/0.13	0.50
$\pi_{22}(\boldsymbol{\theta}(\delta))/\pi_{12}(\boldsymbol{\theta}(\delta))$	0.33/0.33	0.33/0.33	0.33/0.33	0.33/0.33	0.33/0.33	1.00
$\pi_{31}(\boldsymbol{\theta}(\delta))/\pi_{21}(\boldsymbol{\theta}(\delta))$	0.33/0.33	0.26/0.28	0.13/0.17	0.08/0.11	0.06/0.08	0.67
$\pi_{32}(\boldsymbol{\theta}(\delta))/\pi_{22}(\boldsymbol{\theta}(\delta))$	0.33/0.33	0.33/0.33	0.33/0.33	0.33/0.33	0.33/0.33	1.00
$\pi_{41}(\boldsymbol{\theta}(\delta))/\pi_{31}(\boldsymbol{\theta}(\delta))$	0.33/0.33	0.24/0.26	0.11/0.13	0.07/0.08	0.05/0.06	0.75
$\pi_{42}(\boldsymbol{\theta}(\delta))/\pi_{32}(\boldsymbol{\theta}(\delta))$	0.33/0.33	0.33/0.33	0.33/0.33	0.33/0.33	0.33/0.33	1.00

Table 3: Theoretical local odd ratios for the Monte Carlo study.

been considered for some authors, see for instance Cressie et al. (2003) and Martín and Pardo (2012). The cases satisfying the criterion are marked in bold in Table 4, and comprise those values in the abscissa of the plot between the dashed band (the dashed line in the middle represents the nominal size), and we can conclude that we must not consider in our study $T \in \{T_\lambda, S_\lambda\}_{\lambda \in [-1.5, -0.4]}$.

Step 2: We compare all the test statistics obtained in Step 1 with the classical likelihood ratio test ($G^2 = T_0$) as well as the Pearson test statistic ($X^2 = S_1$). To do so, we have calculated the relative local efficiencies

$$\hat{\rho}_T = \hat{\rho}_T(\delta) = \frac{(\hat{\beta}_T(\delta) - \hat{\alpha}_T) - (\hat{\beta}_{T_0}(\delta) - \hat{\alpha}_{T_0})}{\hat{\beta}_{T_0}(\delta) - \hat{\alpha}_{T_0}}, \quad \hat{\rho}_T^* = \hat{\rho}_T^*(\delta) = \frac{(\hat{\beta}_T(\delta) - \hat{\alpha}_T) - (\hat{\beta}_{S_1}(\delta) - \hat{\alpha}_{S_1})}{\hat{\beta}_{S_1}(\delta) - \hat{\alpha}_{S_1}}.$$

It is important to mention that we are comparing the proposed test-statistics with respect to the classical likelihood ratio test ($G^2 = T_0$), which is the only asymptotic test-statistic considered in the literature of hypothesis testing (2a) against (2b), however we are considering also the comparisons with respect to the chi-square test statistic ($X^2 = S_1$) since this is well-known in other ordering types for having good asymptotic performance (see Martín and Balakrishnan (2013) and references therein).

In Figures 1-4 the powers and the relative local efficiencies are summarized. The second rows of the figures represent $\hat{\rho}_T$, while in the third row is plotted $\hat{\rho}_T^*$, on the left it is considered $T = T_\lambda$ and $T = S_\lambda$ on the right.

In all the scenarios a similar pattern is observed when plotting the exact power, $\hat{\beta}_T$, for $\lambda \in (-1, 3)$ since a U shaped curve is obtained. This means that the exact power is higher in the corners of the interval in comparison with the classical likelihood ratio test ($G^2 = T_0$) as well as the classical Pearson test statistic ($X^2 = S_1$), contained in the middle. The likelihood ratio test has very bad performance in relation to the simulated exact size, and we restrict ourselves to $(0, 2]$, taking into account the simulated exact sizes. The Cressie-Read test-statistic ($T_{2/3}$) and the chi-square one ($X^2 = S_1$) have very good performance in regards to the simulated exact size since it is very close to nominal size $\alpha = 0.05$. If we pay attention on the local efficiencies with respect to G^2 and X^2 , $\hat{\rho}_T$ and $\hat{\rho}_T^*$, T_λ and S_λ with λ close to 2 have big values since their powers are greater in comparison with T_λ and S_λ with λ close to 0. For λ close to 2, the values of $\hat{\rho}_T$ are a slightly superior in comparison with $\hat{\rho}_T^*$. Taking into account the plots we conclude that T_{-2} and S_{-2} have clearly the best performance for moderate sample sizes (scenarios 3 and 4) and for small sample sizes (scenarios 1 and 2) the same test-statistics have good performance according to $\hat{\rho}_T$ and $\hat{\rho}_T^*$, however with the Cressie-Read test-statistic ($T_{2/3}$) and the chi-square one ($X^2 = S_1$) a better simulated exact sizes were obtained.

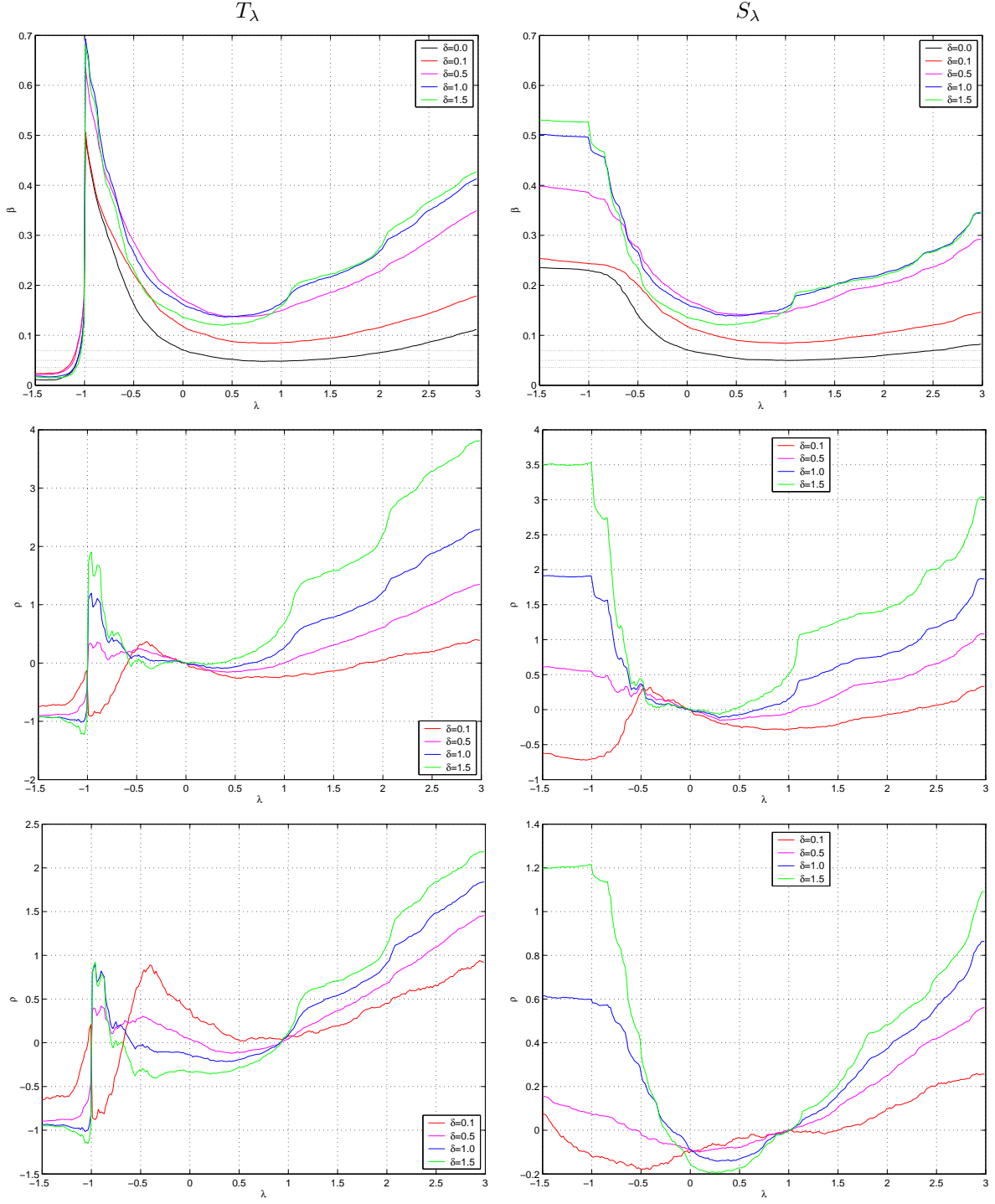


Figure 1: Power and relative local efficiencies for T_λ and S_λ in scenario 1.

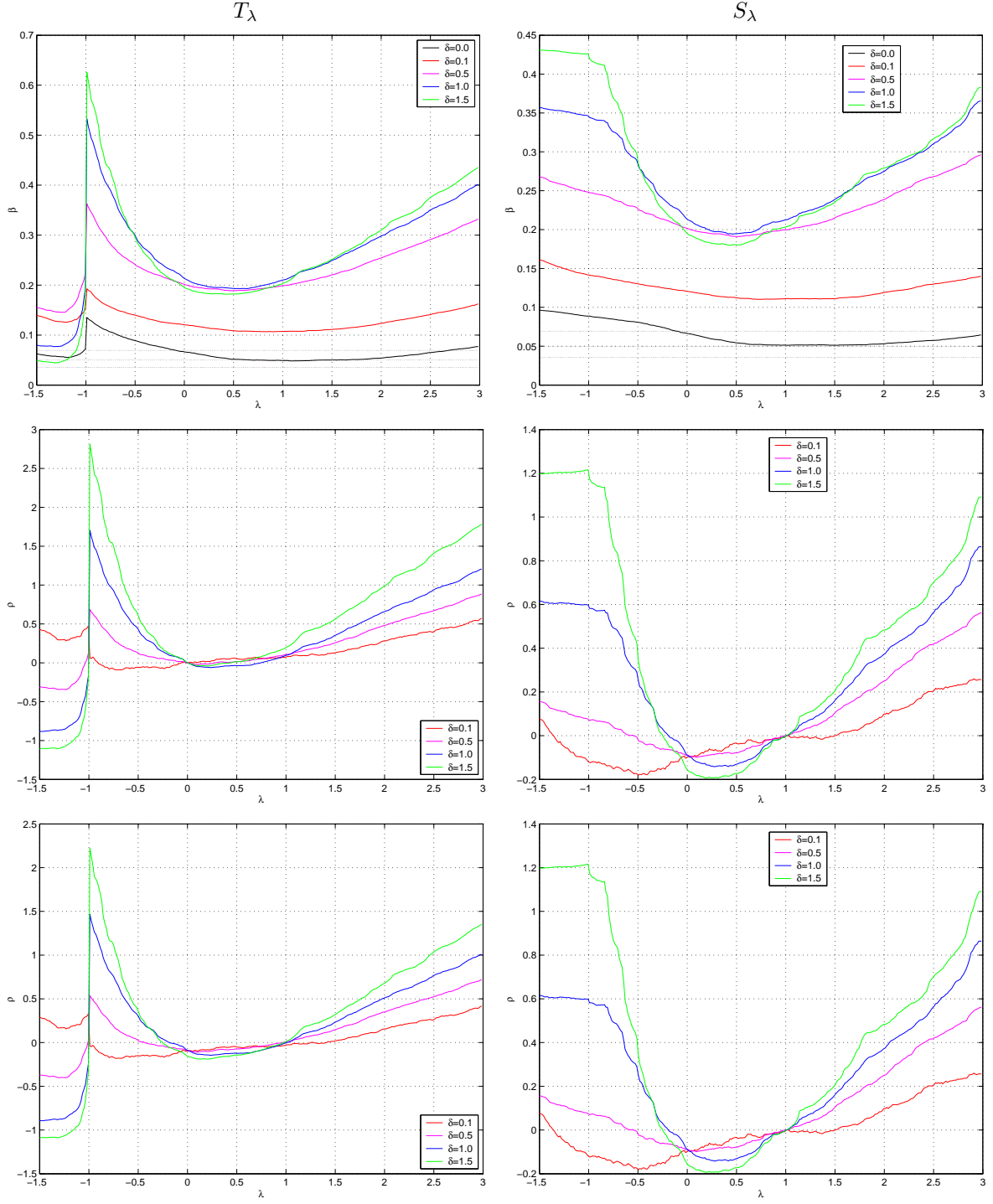


Figure 2: Power and relative local efficiencies for T_λ and S_λ in scenario 2.

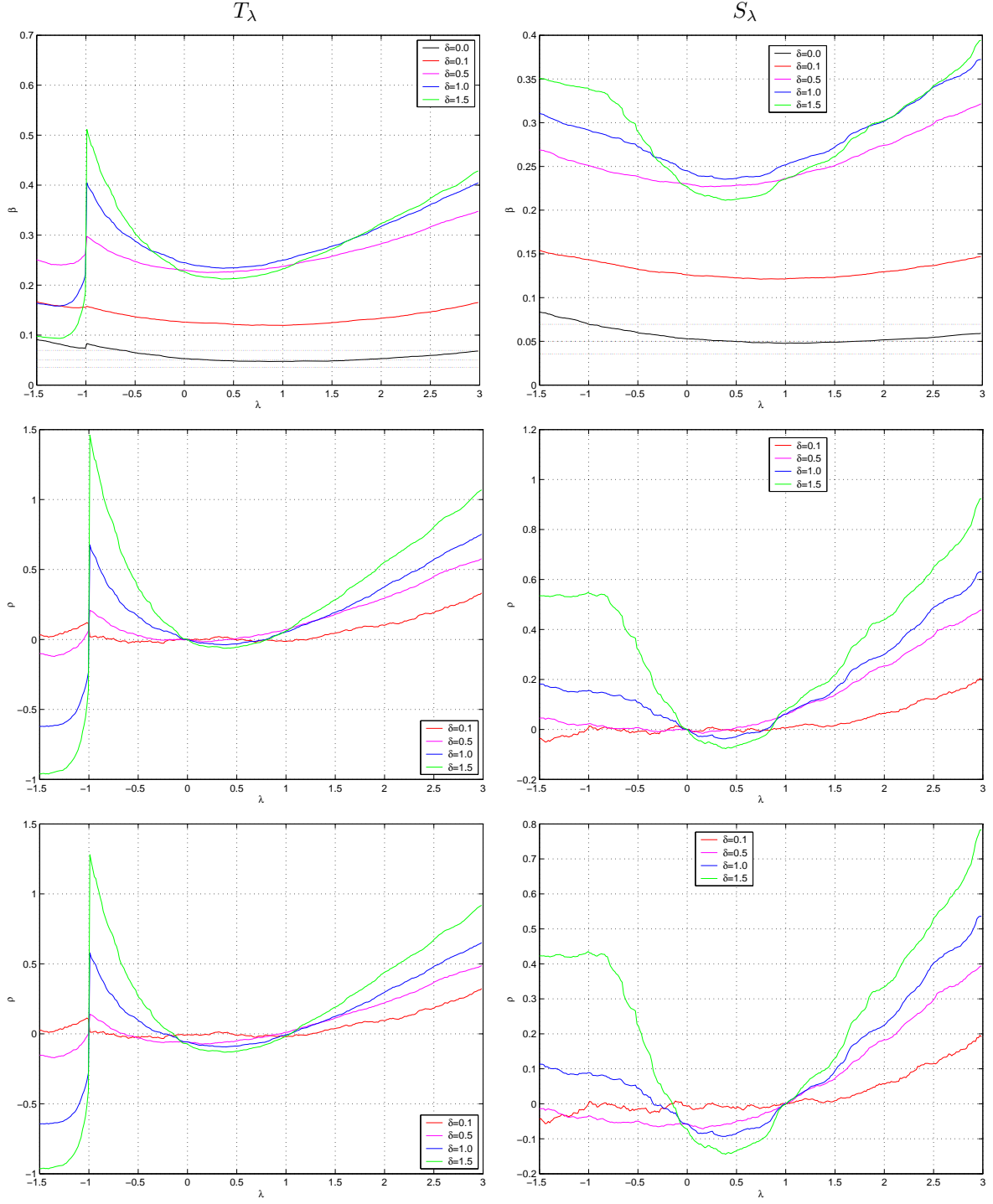


Figure 3: Power and relative local efficiencies for T_λ and S_λ in scenario 3.

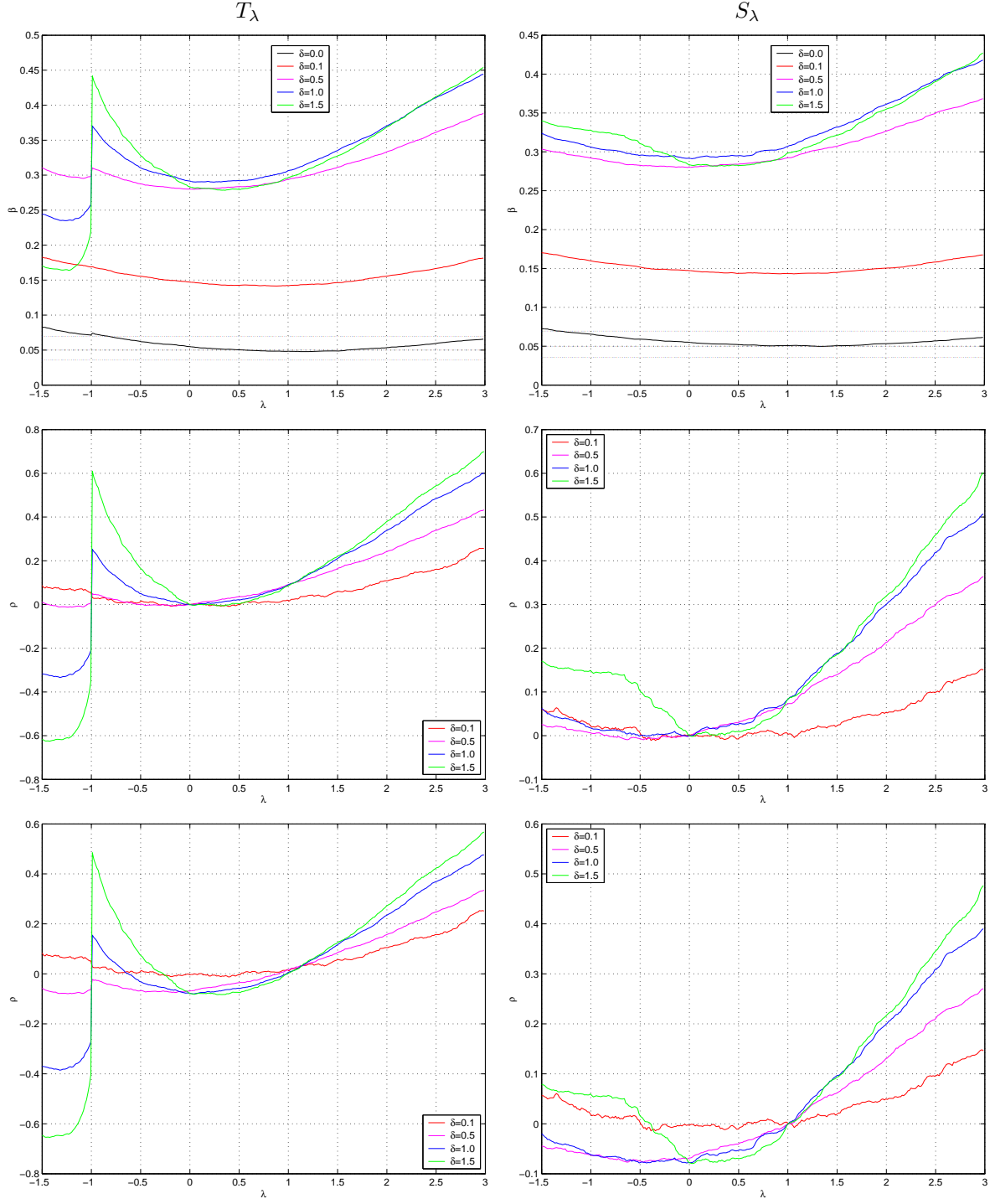


Figure 4: Power and relative local efficiencies for T_λ and S_λ in scenario 4.

Scenario	$\hat{\alpha}_{T_{-1.5}}$	$\hat{\alpha}_{T_{-1}}$	$\hat{\alpha}_{T_{-0.5}}$	$\hat{\alpha}_{T_0}$	$\hat{\alpha}_{T_{\frac{2}{3}}}$	$\hat{\alpha}_{T_1}$	$\hat{\alpha}_{T_{1.5}}$	$\hat{\alpha}_{T_2}$
Scenario 1	0.0111	0.0079	0.1626	0.0706	0.0488	0.0481	0.0533	0.0651
Scenario 2	0.0623	0.0480	0.0888	0.0665	0.0501	0.0492	0.0502	0.0542
Scenario 3	0.0911	0.0702	0.0648	0.0529	0.0477	0.0474	0.0485	0.0531
Scenario 4	0.0827	0.0708	0.0620	0.0550	0.0494	0.0485	0.0487	0.0534
Scenario	$\hat{\alpha}_{S_{-1.5}}$	$\hat{\alpha}_{S_{-1}}$	$\hat{\alpha}_{S_{-0.5}}$	$\hat{\alpha}_{S_0}$	$\hat{\alpha}_{S_{\frac{2}{3}}}$	$\hat{\alpha}_{S_1}$	$\hat{\alpha}_{S_{1.5}}$	$\hat{\alpha}_{S_2}$
Scenario 1	0.2356	0.2299	0.1409	0.0706	0.0514	0.0498	0.0527	0.0599
Scenario 2	0.0966	0.0887	0.0809	0.0665	0.0521	0.0514	0.0512	0.0532
Scenario 3	0.0835	0.0697	0.0597	0.0529	0.0485	0.0479	0.0490	0.0516
Scenario 4	0.0726	0.0654	0.0590	0.0550	0.0516	0.0508	0.0505	0.0533

Table 4: $\hat{\alpha}_T$, for $T \in \{T_\lambda, S_\lambda\}_{\lambda \in \Lambda}$ in the four scenarios.

7 Summary and conclusion

We have proposed and studied two new families of test-statistics, useful for testing if there exists homogeneity in I independent multinomial samples or on the contrary, likelihood ordering. Their asymptotic chi-bar distribution is common, with weights easy to be estimated, using the matrix $\mathbf{H}(\hat{\theta})$, and have a simple interpretation in terms of log-linear modeling. Two algorithms provide the procedure for computing the estimation of the weights and asymptotic p -values of the test-statistics. In the literature of likelihood ratio ordering, using asymptotic techniques, the likelihood ratio test has solely been considered. The simulation study shows that such a test-statistic has a poor performance for small and moderate sample sizes and we have seen that it is much better using other test-statistics such as T_λ and S_λ with $\lambda = 2$.

References

- [1] BARLOW, R. E., BARTHOLOMEW, D. J. AND BRUNK, H.D. (1972). *Statistical inference under order restrictions*. Wiley.
- [2] BAPAT, R. B., KOCHAR, S. C. (1994). On Likelihood-Ratio Ordering of Order Statistics. *Matrix algebra and its applications*, **199**, 281–291.
- [3] BAZARAA, M. S., SHERALI, H. D. AND SHETTY, C. M. (2006). *Nonlinear Programming: Theory and Algorithms* (3rd Edition). John Wiley and Sons.
- [4] CHRISTENSEN, R. (1997). *Log-linear models and logistic regression*. Springer.
- [5] CRESSIE, N., PARDO, L. AND PARDO, M.C. (2003). Size and power considerations for testing loglinear models using ϕ -divergence test statistics. *Statistica Sinica*, **13**, 550-570.
- [6] DARDANONI, V. AND FORCINA, A. (1998). A Unified Approach to Likelihood Inference on Stochastic Orderings in a Nonparametric Context. *Journal of Americal Association*, **93**, 1112–1123.
- [7] DAVIDOV, O., FOKIANOS, K., ILIOPOULOS, G. (2010). Order-Restricted Semiparametric Inference for the Power Bias Model. *Biometrics* **66**, 549-557.
- [8] DYKSTRA, R. L., KOCBAR, S. AND ROBERTSON, T. (1995). Inference for Likelihood Ratio Ordering in the Two-Sample Problem. *Journal of the American Statistical Association*, **90**, 1034-1040.
- [9] HARVILLE, D. A. (2008). *Matrix algebra from a statistician's perspective*. Springer.

- [10] FERGUSON, T. S. (1996). *A Course in Large Sample Theory*. Chapman & Hall.
- [11] KUDÔ, A. (1963). A multivariate analogue of the one-sided test. *Biometrika*, **50**, 403-418.
- [12] LANG, J. B. (1996). On the Comparison of Multinomial and Poisson Log-Linear Models. *Journal of the Royal Statistical Society Series B*, **58**, 253-266.
- [13] MARTIN, N. AND BALAKRISHNAN, N. (2013). Hypothesis testing in a generic nesting framework for general distributions. *Journal of Multivariate Analysis*, **118**, 1-23.
- [14] MARTIN, N. AND PARDO, L. (2008) New families of estimators and test statistics in log-linear models. *Journal of Multivariate Analysis*, **99**(8), 1590-1609.
- [15] MARTIN, N. AND PARDO, L. (2012). Poisson-loglinear modeling with linear constraints on the expected cell frequencies. *Sankhya B*, **74**(2), 238-267.
- [16] PARDO, L. (2006). *Statistical Inference Based on Divergence Measures*. Statistics: series of Textbooks and Monographs. Chapman & Hall / CRC.
- [17] SEN, P. K., SINGER, J. M. AND PEDROSO DE LIMA, A. C. (2010). *From Finite Sample to Asymptotic Methods in Statistics*. Cambridge University Press.
- [18] SHAPIRO, A. (1985). Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures Under Inequality Constraints. *Biometrika*, **72**, 133-144.
- [19] SHAPIRO, A. (1988). Toward a Unified Theory of Inequality Constrained Testing in Multivariate Analysis. *International Statistical Review*, **56**, 49-62.
- [20] SILVAPULLE, M. J. AND SEN., P. K. (2005). *Constrained statistical inference. Inequality, order, and shape restrictions*. Wiley Series in Probability and Statistics. Wiley-Interscience (John Wiley & Sons).
- [21] Wang, Y. (1996). A likelihood ratio test against Stochastic ordering in several populations. *Journal of Americal Association*, **91**, 1676-1683.

A Appendix

Suppose we are interested in testing

$$H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{0}_{(I-1)(J-1)} \quad \text{vs} \quad H_1 : \mathbf{R}(S)\boldsymbol{\theta} = \mathbf{0}_{\text{card}(S)} \quad \text{and} \quad \mathbf{R}\boldsymbol{\theta} \neq \mathbf{0}_{(I-1)(J-1)}. \quad (40)$$

Under H_0 , the parameter space is $\Theta_0 = \{\boldsymbol{\theta} \in \mathbb{R}^{I(J-1)} : \mathbf{R}\boldsymbol{\theta} = \mathbf{0}_{(I-1)(J-1)}\}$ and the maximum likelihood estimator of $\boldsymbol{\theta}$ in Θ_0 is $\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta_0} \ell(\mathbf{N}; \boldsymbol{\theta})$. Under the alternative hypothesis the parameter space is $\Theta(S) - \Theta_0$, where $\Theta(S) = \{\boldsymbol{\theta} \in \mathbb{R}^{I(J-1)} : \mathbf{R}(S)\boldsymbol{\theta} = \mathbf{0}_{(I-1)(J-1)}\}$, that is, under both hypotheses, H_0 and H_1 , the parameter space is $\Theta(S) = \{\boldsymbol{\theta} \in \mathbb{R}^{I(J-1)} : \mathbf{R}(S)\boldsymbol{\theta} = \mathbf{0}_{(I-1)(J-1)}\}$ and the maximum likelihood estimator of $\boldsymbol{\theta}$ in $\Theta(S)$ is $\hat{\boldsymbol{\theta}}(S) = \arg \max_{\boldsymbol{\theta} \in \Theta(S)} \ell(\mathbf{N}; \boldsymbol{\theta})$. By following the same idea we used for building test-statistics (17)-(18) we shall consider two family of test-statistics based on ϕ -divergence measures,

$$T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}(S)), \mathbf{p}(\hat{\boldsymbol{\theta}})) = \frac{2n}{\phi'(1)} (d_\phi(\bar{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}})) - d_\phi(\bar{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}(S)))) \quad (41)$$

and

$$S_\phi(\mathbf{p}(\hat{\boldsymbol{\theta}}(S)), \mathbf{p}(\hat{\boldsymbol{\theta}})) = \frac{2n}{\phi'(1)} d_\phi(\mathbf{p}(\hat{\boldsymbol{\theta}}(S)), \mathbf{p}(\hat{\boldsymbol{\theta}})). \quad (42)$$

A.1 Proposition

Under H_0 ,

$$S_\phi(\mathbf{p}(\widehat{\boldsymbol{\theta}}(S)), \mathbf{p}(\widehat{\boldsymbol{\theta}})) = T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\widehat{\boldsymbol{\theta}}(S)), \mathbf{p}(\widehat{\boldsymbol{\theta}})) + o_p(1), \quad (43)$$

the asymptotic distribution of (41) and (42) is χ_{df}^2 with $df = (I-1)(J-1) - \text{card}(S)$.

Proof. The second order Taylor expansion of function $d_\phi(\boldsymbol{\theta}) = d_\phi(\mathbf{p}(\boldsymbol{\theta}), \mathbf{p}(\widehat{\boldsymbol{\theta}}))$ about $\widehat{\boldsymbol{\theta}}$ is

$$d_\phi(\boldsymbol{\theta}) = d_\phi(\widehat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^T \frac{\partial}{\partial \boldsymbol{\theta}} d_\phi(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} + \frac{1}{2} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^T \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} d_\phi(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) + o\left(\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|^2\right), \quad (44)$$

where

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} d_\phi(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} &= \mathbf{0}_{(I-1)(J-1)}, \\ \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} d_\phi(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} &= \phi''(1) \mathcal{I}_F^{(n_1, \dots, n_I)}(\widehat{\boldsymbol{\theta}}), \end{aligned}$$

and $\mathcal{I}_F^{(n_1, n_2)}(\boldsymbol{\theta})$ was defined at the beginning of Section 4. Let $\bar{\boldsymbol{\theta}}$ such that $\bar{\mathbf{p}} = \mathbf{p}(\bar{\boldsymbol{\theta}})$, where $\mathbf{p}(\bar{\boldsymbol{\theta}}) = \exp\{\mathbf{1}_{IJ}\bar{u} + \mathbf{W}_1\bar{\boldsymbol{\theta}}_1 + \mathbf{W}\bar{\boldsymbol{\theta}}\}$, with $\bar{u} = u(\bar{\boldsymbol{\theta}})$, $\bar{\boldsymbol{\theta}}_{1(i)} = \boldsymbol{\theta}_{1(i)}(\bar{\boldsymbol{\theta}})$, is the the saturated log-linear model. In particular, for $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$ we have

$$d_\phi(\mathbf{p}(\bar{\boldsymbol{\theta}}), \mathbf{p}(\widehat{\boldsymbol{\theta}})) = \frac{\phi''(1)}{2} (\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}})^T \mathcal{I}_F^{(n_1, \dots, n_I)}(\widehat{\boldsymbol{\theta}}) (\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}) + o\left(\|\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}\|^2\right).$$

In a similar way it is obtained

$$d_\phi(\mathbf{p}(\bar{\boldsymbol{\theta}}), \mathbf{p}(\widehat{\boldsymbol{\theta}}(S))) = \frac{\phi''(1)}{2} (\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(S))^T \mathcal{I}_F^{(n_1, \dots, n_I)}(\widehat{\boldsymbol{\theta}}(S)) (\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(S)) + o\left(\|\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(S)\|^2\right).$$

Multiplying both sides of the equality by $\frac{2n}{\phi''(1)}$ and taking the difference in both sides of the equality

$$\begin{aligned} T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\widehat{\boldsymbol{\theta}}(S)), \mathbf{p}(\widehat{\boldsymbol{\theta}})) &= \frac{2n}{\phi''(1)} \left(d_\phi(\mathbf{p}(\bar{\boldsymbol{\theta}}), \mathbf{p}(\widehat{\boldsymbol{\theta}})) - d_\phi(\mathbf{p}(\bar{\boldsymbol{\theta}}), \mathbf{p}(\widehat{\boldsymbol{\theta}}(S))) \right) \\ &= \sqrt{n}(\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}})^T \mathcal{I}_F^{(n_1, \dots, n_I)}(\widehat{\boldsymbol{\theta}}) \sqrt{n}(\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}) + o\left(\left\|\sqrt{n}(\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}})\right\|^2\right) \\ &\quad - \sqrt{n}(\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(S))^T \mathcal{I}_F^{(n_1, \dots, n_I)}(\widehat{\boldsymbol{\theta}}(S)) \sqrt{n}(\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(S)) + o\left(\left\|\sqrt{n}(\bar{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}(S))\right\|^2\right). \end{aligned}$$

Now we are going to generalize the three types of estimators by $\widehat{\boldsymbol{\theta}}(\bullet)$, understanding that for $\bullet = \emptyset$, $\widehat{\boldsymbol{\theta}}(\emptyset) = \bar{\boldsymbol{\theta}}$, $\mathbf{R}(\emptyset) = \mathbf{0}_{(I-1)(J-1) \times (I-1)(J-1)}$, for $\bullet = E$, $\widehat{\boldsymbol{\theta}}(E) = \widehat{\boldsymbol{\theta}}$, $\mathbf{R}(E) = \mathbf{R}$, and $\bullet = S$, $\widehat{\boldsymbol{\theta}}(S)$ and $\mathbf{R}(S)$ as originally defined. It is well-known that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}(\bullet) - \boldsymbol{\theta}_0) = \mathbf{P}(\boldsymbol{\theta}_0, \bullet) \frac{1}{\sqrt{n}} \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\mathbf{N}; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + o_p(\mathbf{1}_k), \quad (45)$$

where $\boldsymbol{\theta}_0$ is the true and unknown value of the parameter,

$$\mathbf{P}(\boldsymbol{\theta}_0, \bullet) = \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) - \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{R}^T(\bullet) \left(\mathbf{R}(\bullet) \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{R}^T(\bullet) \right)^{-1} \mathbf{R}(\bullet) \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0),$$

is the variance covariance matrix of $\widehat{\boldsymbol{\theta}}(\bullet)$, and $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\mathbf{N}; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_k, \mathcal{I}_F(\boldsymbol{\theta}_0))$ by the Central Limit Theorem. We shall denote

$$\mathbf{P}(\boldsymbol{\theta}_0) = \mathbf{P}(\boldsymbol{\theta}_0, E) = \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) - \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{R}^T \left(\mathbf{R} \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{R}^T \right)^{-1} \mathbf{R} \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0).$$

Taking the differences of both sides of the equality in (45) with cases $\bullet = \emptyset$ and $\bullet = E$, we obtain

$$\sqrt{n}(\bar{\theta} - \hat{\theta}) = (\mathcal{I}_F^{-1}(\theta_0) - \mathbf{P}(\theta_0)) \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \ell(\mathbf{N}; \theta) \Big|_{\theta=\theta_0} + o_p(\mathbf{1}_k), \quad (46)$$

with cases $\bullet = \emptyset$ and $\bullet = S$,

$$\sqrt{n}(\bar{\theta} - \hat{\theta}(S)) = (\mathcal{I}_F^{-1}(\theta_0) - \mathbf{P}(\theta_0, S)) \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \ell(\mathbf{N}; \theta) \Big|_{\theta=\theta_0} + o_p(\mathbf{1}_k), \quad (47)$$

and taking into account $\mathcal{I}_F(\hat{\theta}) \xrightarrow[n \rightarrow \infty]{P} \mathcal{I}_F(\theta_0)$,

$$\begin{aligned} & T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\hat{\theta}(S)), \mathbf{p}(\hat{\theta})) \\ &= \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta^T} \ell(\mathbf{N}; \theta) \Big|_{\theta=\theta_0} (\mathbf{P}(\theta_0, S) - \mathbf{P}(\theta_0))^T \mathcal{I}_F(\theta_0) (\mathbf{P}(\theta_0, S) - \mathbf{P}(\theta_0)) \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \ell(\mathbf{N}; \theta) \Big|_{\theta=\theta_0} + o_p(1) \\ &= \mathbf{Y}^T \mathbf{Y} + o_p(1), \end{aligned} \quad (48)$$

where

$$\mathbf{Y} = \mathbf{A}(\theta_0) (\mathbf{P}(\theta_0, S) - \mathbf{P}(\theta_0)) \mathbf{A}(\theta_0)^T \mathbf{Z},$$

with $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_{(I-1)(J-1)}, \mathbf{I}_{(I-1)(J-1)})$ and $\mathbf{A}(\theta_0)$ is the Cholesky's factorization matrix for a non singular matrix such a Fisher information matrix, that is $\mathcal{I}_F(\theta_0) = \mathbf{A}(\theta_0)^T \mathbf{A}(\theta_0)$. In other words

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{0}_k, \mathbf{A}(\theta_0) (\mathbf{P}(\theta_0, S) - \mathbf{P}(\theta_0)) \mathbf{A}(\theta_0)^T),$$

where the variance covariance matrix is idempotent and symmetric. Following Lemma 3 in Ferguson (1996, page 57), $\mathbf{A}(\theta_0) (\mathbf{P}(\theta_0, S) - \mathbf{P}(\theta_0)) \mathbf{A}(\theta_0)^T$ is idempotent and symmetric, if only if $T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\hat{\theta}(S)), \mathbf{p}(\hat{\theta}))$ is a chi-square random variable with degrees of freedom

$$df = \text{rank}(\mathbf{A}(\theta_0) (\mathbf{P}(\theta_0, S) - \mathbf{P}(\theta_0)) \mathbf{A}(\theta_0)^T) = \text{trace}(\mathbf{A}(\theta_0) (\mathbf{P}(\theta_0, S) - \mathbf{P}(\theta_0)) \mathbf{A}(\theta_0)^T).$$

Since

$$(\mathbf{P}(\theta_0, S) - \mathbf{P}(\theta_0))^T \mathcal{I}_F(\theta_0) (\mathbf{P}(\theta_0, S) - \mathbf{P}(\theta_0)) = \mathbf{P}(\theta_0, S) - \mathbf{P}(\theta_0),$$

the condition is reached. The effective degrees of freedom are given by

$$\begin{aligned} df &= \text{trace}(\mathbf{P}(\theta_0, S) \mathbf{A}(\theta_0)^T \mathbf{A}(\theta_0)) - \text{trace}(\mathbf{P}(\theta_0) \mathbf{A}(\theta_0)^T \mathbf{A}(\theta_0)) = \text{trace}(\mathbf{P}(\theta_0, S) \mathcal{I}_F(\theta_0)) - \text{trace}(\mathbf{P}(\theta_0) \mathcal{I}_F(\theta_0)) \\ &= \text{trace}(-\left(\mathbf{R}(S) \mathcal{I}_F^{-1}(\theta_0) \mathbf{R}^T(S)\right)^{-1} \mathbf{R}(S) \mathcal{I}_F^{-1}(\theta_0) \mathbf{R}^T(S)) \\ &\quad - \text{trace}(-\left(\mathbf{R} \mathcal{I}_F^{-1}(\theta_0) \mathbf{R}^T\right)^{-1} \mathbf{R} \mathcal{I}_F^{-1}(\theta_0) \mathbf{R}^T) \\ &= (I-1)(J-1) - \text{card}(S). \end{aligned}$$

Regarding the other test-statistic $S_\phi(\mathbf{p}(\hat{\theta}(S)), \mathbf{p}(\hat{\theta}))$, observe that if we take (44), in particular for $\theta = \hat{\theta}(S)$ it is obtained

$$d_\phi(\hat{\theta}(S)) = \frac{\phi''(1)}{2} (\hat{\theta}(S) - \hat{\theta})^T \mathcal{I}_F(\hat{\theta}) (\hat{\theta}(S) - \hat{\theta}) + o\left(\|\hat{\theta}(S) - \hat{\theta}\|^2\right).$$

In addition, (46)–(47) is

$$\sqrt{n}(\hat{\theta}(S) - \hat{\theta}) = (\mathbf{P}(\theta_0, S) - \mathbf{P}(\theta_0)) \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \ell(\mathbf{N}; \theta) \Big|_{\theta=\theta_0} + o_p(\mathbf{1}_k),$$

and taking into account $\mathcal{I}_F(\hat{\theta}) \xrightarrow[n \rightarrow \infty]{P} \mathcal{I}_F(\theta_0)$ and (48), it follows (43), which means from Slutsky's Theorem that both test-statistics have the same asymptotic distribution. ■

A.2 Lemma

Let \mathbf{Y} be a k -dimensional random variable with normal distribution $\mathcal{N}(\mathbf{0}_k, \mathbf{I})$ with \mathbf{Q} being a projection matrix, that is idempotent and symmetric, and let fixed k -dimensional vectors \mathbf{d}_i such that for them either $\mathbf{Q}\mathbf{d}_i = \mathbf{0}_k$ or $\mathbf{Q}\mathbf{d}_i = \mathbf{d}_i$, $i = 1, \dots, k$, is true. Then $\left(\mathbf{Y}^T \mathbf{Q} \mathbf{Y} \mid \mathbf{d}_i^T \mathbf{Y} \geq 0, i = 1, \dots, k\right) \sim \chi_{df}^2$, where $df = \text{rank}(\mathbf{Q})$.

Proof. This result can be found in several sources, for instance in Kudô (1963, page 414), Barlow et al. (1972, page 128) and Shapiro (1985, page 139). ■

A.3 Proof of Theorem 2

We shall perform the proof for $S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$. It suppose that it is true $\mathbf{R}\boldsymbol{\theta} \geq \mathbf{0}_{(I-1)(J-1)}$ and we want to test $\mathbf{R}\boldsymbol{\theta} = \mathbf{0}_{(I-1)(J-1)}$ (H_0). It is clear that if H_0 is not true is because there exists some index $i \in E$ such that $\mathbf{R}(\{i\})\boldsymbol{\theta} > 0$. Let us consider the family of all possible subsets in E , denoted by $\mathcal{F}(E)$, then we shall specify more thoroughly $\tilde{\boldsymbol{\theta}}$ by $\tilde{\boldsymbol{\theta}}(S)$ when there exists $S \in \mathcal{F}(E)$ such that

$$\mathbf{R}(S)\tilde{\boldsymbol{\theta}} = \mathbf{0}_{\text{card}(S)} \quad \text{and} \quad \mathbf{R}(S^C)\tilde{\boldsymbol{\theta}} > \mathbf{0}_{(I-1)(J-1)-\text{card}(S)}.$$

It is clear that for a sample $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}(S)$ can be true only for a unique set of indices $S \in \mathcal{F}(E)$, and thus by applying the Theorem of Total Probability

$$\Pr\left(S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \leq x\right) = \sum_{S \in \mathcal{F}(E)} \Pr\left(S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \leq x, \tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}(S)\right).$$

From the Karush-Khun-Tucker necessary conditions (see for instance Theorem 4.2.13 in Bazaraa et al. (2006)) to solve the optimization problem $\max \ell(\mathbf{N}; \boldsymbol{\theta})$ s.t. $\mathbf{R}\boldsymbol{\theta} \geq \mathbf{0}_{(I-1)(J-1)}$, associated with $\tilde{\boldsymbol{\theta}}$,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ell(\mathbf{N}; \boldsymbol{\theta}) + \sum_{i=1}^{(I-1)(J-1)} \lambda_i \mathbf{R}^T(\{i\}) = 0, \quad i = 1, \dots, (I-1)(J-1), \quad (49a)$$

$$\lambda_i \mathbf{R}(\{i\})\boldsymbol{\theta} = 0, \quad i = 1, \dots, (I-1)(J-1), \quad (49b)$$

$$\lambda_i \leq 0, \quad i = 1, \dots, (I-1)(J-1), \quad (49c)$$

the only conditions which characterize the MLE $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}(S)$ with a specific $S \in \mathcal{F}(E)$, are the complementary slackness conditions $\mathbf{R}(\{i\})\boldsymbol{\theta} > 0$, for $i \in S$ and $\lambda_i < 0$, for $i \in S^C$, since $\frac{\partial}{\partial \boldsymbol{\theta}} \ell(\mathbf{N}; \boldsymbol{\theta}) + \lambda_i \mathbf{R}^T(\{i\}) = 0$, $i = 1, \dots, (I-1)(J-1)$, $\mathbf{R}(\{i\})\boldsymbol{\theta} = 0$, for $i \in S^C$ and $\lambda_i = 0$, for $i \in S$ are redundant conditions once we know that the Karush-Khun-Tucker necessary conditions are true for all the possible sets $S \in \mathcal{F}(E)$ which define $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}(S)$. For this reason we can consider

$$\begin{aligned} & \Pr\left(S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \leq x, \tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}(S)\right) = \\ & \Pr\left(S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \leq x, \tilde{\boldsymbol{\lambda}}(S) < \mathbf{0}_{\text{card}(S)}, \mathbf{R}(S^C)\tilde{\boldsymbol{\theta}}(S) > \mathbf{0}_{(I-1)(J-1)-\text{card}(S)}\right), \end{aligned}$$

where $\tilde{\boldsymbol{\lambda}}(S)$ is the vector of the vector of Karush-Khun-Tucker multipliers associated with estimator $\tilde{\boldsymbol{\theta}}(S)$. Furthermore, under H_0 , $\mathbf{R}\tilde{\boldsymbol{\theta}}(S) = \mathbf{R}\tilde{\boldsymbol{\theta}}(S) - \mathbf{R}\boldsymbol{\theta}_0$, because $\mathbf{R}\boldsymbol{\theta}_0 = \mathbf{0}_{(I-1)(J-1)}$, hence

$$\Pr\left(S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \leq x\right) = \sum_{S \in \mathcal{F}(E)} \Pr\left(S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \leq x, \tilde{\boldsymbol{\lambda}}(S) < \mathbf{0}_{\text{card}(S)}, \mathbf{R}(S^C)\tilde{\boldsymbol{\theta}}(S) - \mathbf{R}(S^C)\boldsymbol{\theta}_0 > \mathbf{0}_{\text{card}(S^C)}\right),$$

where $\text{card}(S^C) = (I-1)(J-1) - \text{card}(S)$. On the other hand, (49a) and (49b) are also true for $(\hat{\boldsymbol{\theta}}^T(S), \hat{\boldsymbol{\lambda}}^T(S))^T$ according to the Lagrange multipliers method. Hence, $\tilde{\boldsymbol{\theta}}(S) = \hat{\boldsymbol{\theta}}(S)$ and $\tilde{\boldsymbol{\lambda}}(S) = \hat{\boldsymbol{\lambda}}(S)$. It follows that:

- under $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(S)$, $S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = S_\phi(\mathbf{p}(\hat{\boldsymbol{\theta}}(S)), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ and taking into account the Proposition given in Section A.1

$$\begin{aligned} S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) &= T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}(S)), \mathbf{p}(\hat{\boldsymbol{\theta}})) + o_p(1) \\ &= \left(\mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{P}(\boldsymbol{\theta}_0, S) - \mathbf{P}(\boldsymbol{\theta}_0)) \mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{Z} \right)^T \left(\mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{P}(\boldsymbol{\theta}_0, S) - \mathbf{P}(\boldsymbol{\theta}_0)) \mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{Z} \right) + o_p(1), \\ &= \mathbf{Z}^T \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{P}(\boldsymbol{\theta}_0, S) - \mathbf{P}(\boldsymbol{\theta}_0)) \mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{Z} + o_p(1). \end{aligned}$$

where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_k, \mathbf{I}_k)$.

- under $\tilde{\boldsymbol{\lambda}}(S) = \hat{\boldsymbol{\lambda}}(S)$ and from Sen et al. (2010, page 267 formula (8.6.28))

$$\begin{aligned} \frac{1}{\sqrt{n}} \tilde{\boldsymbol{\lambda}}(S) &= \sqrt{n} \mathbf{Q}^T(\boldsymbol{\theta}_0, S) \frac{1}{\sqrt{n}} \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\mathbf{N}; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + o_p(\mathbf{1}_{\text{card}(S)}) \\ &= \mathbf{Q}^T(\boldsymbol{\theta}_0, S) \mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{Z} + o_p(\mathbf{1}_{\text{card}(S)}), \end{aligned}$$

where

$$\mathbf{Q}(\boldsymbol{\theta}_0, S) = -\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{R}^T(S) \mathbf{L}(\boldsymbol{\theta}_0, S) \left(\mathbf{R}(S) \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{R}^T(S) \right)^{-1};$$

- under $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(S)$ and from (45)

$$\begin{aligned} \sqrt{n} \left(\mathbf{R}(S^C) \tilde{\boldsymbol{\theta}}(S) - \mathbf{R}(S^C) \boldsymbol{\theta}_0 \right) &= \sqrt{n} \mathbf{R}(S^C) \mathbf{P}(\boldsymbol{\theta}_0, S) \frac{1}{\sqrt{n}} \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\mathbf{N}; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + o_p(\mathbf{1}_{\text{card}(S^C)}) \\ &= \mathbf{R}(S^C) \mathbf{P}(\boldsymbol{\theta}_0, S) \mathbf{A}(\boldsymbol{\theta}_0)^T \mathbf{Z} + o_p(\mathbf{1}_{\text{card}(S^C)}). \end{aligned}$$

That is,

$$\begin{aligned} \lim_{n \rightarrow \infty} \Pr \left(S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \leq x \right) &= \sum_{S \in \mathcal{F}(E)} \Pr \left(\mathbf{Z}_3^T(S) \mathbf{Z}_3(S) \leq x, \mathbf{Z}_1(S) \geq \mathbf{0}_{\text{card}(S)}, \mathbf{Z}_2(S) \geq \mathbf{0}_{\text{card}(S^C)} \right) \\ &= \sum_{S \in \mathcal{F}(E)} \Pr \left(\mathbf{Z}_3^T(S) \mathbf{Z}_3(S) \leq x \mid \mathbf{Z}_1(S) \geq \mathbf{0}_{\text{card}(S)}, \mathbf{Z}_2(S) \geq \mathbf{0}_{\text{card}(S^C)} \right) \Pr \left(\mathbf{Z}_1(S) \geq \mathbf{0}_{\text{card}(S)}, \mathbf{Z}_2(S) \geq \mathbf{0}_{\text{card}(S^C)} \right) \\ &= \sum_{S \in \mathcal{F}(E)} \Pr \left(\mathbf{Z}_3^T(S) \mathbf{Z}_3(S) \leq x \mid \left(\mathbf{Z}_1^T(S), \mathbf{Z}_2^T(S) \right)^T \geq \mathbf{0}_{(I-1)(J-1)} \right) \Pr \left(\mathbf{Z}_1(S) \geq \mathbf{0}_{\text{card}(S)}, \mathbf{Z}_2(S) \geq \mathbf{0}_{\text{card}(S^C)} \right), \end{aligned}$$

where

$$\begin{aligned} \mathbf{Z}_3(S) &= \mathbf{M}_3(\boldsymbol{\theta}_0, S) \mathbf{Z}, & \mathbf{M}_3(\boldsymbol{\theta}_0, S) &= \mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{P}(\boldsymbol{\theta}_0, S) - \mathbf{P}(\boldsymbol{\theta}_0)) \mathbf{A}(\boldsymbol{\theta}_0)^T, \\ \mathbf{Z}_1(S) &= \mathbf{M}_1(\boldsymbol{\theta}_0, S) \mathbf{Z}, & \mathbf{M}_1(\boldsymbol{\theta}_0, S) &= -\mathbf{Q}^T(\boldsymbol{\theta}_0, S) \mathbf{A}(\boldsymbol{\theta}_0)^T, \\ \mathbf{Z}_2(S) &= \mathbf{M}_2(\boldsymbol{\theta}_0, S) \mathbf{Z}, & \mathbf{M}_2(\boldsymbol{\theta}_0, S) &= \mathbf{R}(S^C) \mathbf{P}(\boldsymbol{\theta}_0, S) \mathbf{A}(\boldsymbol{\theta}_0)^T. \end{aligned}$$

Taking into account that $\mathbf{M}_3(\boldsymbol{\theta}_0, S) \mathbf{M}_2^T(\boldsymbol{\theta}_0, S) = \mathbf{M}_2^T(\boldsymbol{\theta}_0, S)$ and $\mathbf{M}_3(\boldsymbol{\theta}_0, S) \mathbf{M}_1^T(\boldsymbol{\theta}_0, S) = \mathbf{0}_{(I-1)(J-1) \times \text{card}(S)}$, by applying the lemma given in Section A.2

$$\Pr \left(\mathbf{Z}_3^T(S) \mathbf{Z}_3(S) \leq x \mid \left(\mathbf{Z}_1^T(S), \mathbf{Z}_2^T(S) \right)^T \geq \mathbf{0}_{(I-1)(J-1)} \right) = \Pr \left(\chi_{df}^2 \leq x \right)$$

where

$$\begin{aligned} df &= \text{rank} \left(\mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{P}(\boldsymbol{\theta}_0, S) - \mathbf{P}(\boldsymbol{\theta}_0)) \mathbf{A}(\boldsymbol{\theta}_0)^T \right) = \text{trace} \left(\mathbf{A}(\boldsymbol{\theta}_0) (\mathbf{P}(\boldsymbol{\theta}_0, S) - \mathbf{P}(\boldsymbol{\theta}_0)) \mathbf{A}(\boldsymbol{\theta}_0)^T \right) \\ &= (I-1)(J-1) - \text{card}(S). \end{aligned}$$

Finally,

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \Pr \left(S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \leq x \right) \\
&= \sum_{S \in \mathcal{F}(E)} \Pr \left(\chi_{(I-1)(J-1) - \text{card}(S)}^2 \leq x \right) \Pr \left(\mathbf{Z}_1(S) \geq \mathbf{0}_{\text{card}(S)}, \mathbf{Z}_2(S) \geq \mathbf{0}_{\text{card}(S^C)} \right) \\
&= \sum_{j=0}^{(I-1)(J-1)} \Pr \left(\chi_{(I-1)(J-1)-j}^2 \leq x \right) \sum_{S \in \mathcal{F}(E), \text{card}(S)=j} \Pr \left(\mathbf{Z}_1(S) \geq \mathbf{0}_{\text{card}(S)}, \mathbf{Z}_2(S) \geq \mathbf{0}_{\text{card}(S^C)} \right),
\end{aligned}$$

and since $\mathbf{Q}^T(\boldsymbol{\theta}_0, S) \mathcal{I}_F(\boldsymbol{\theta}_0) \mathbf{P}(\boldsymbol{\theta}_0, S) = \mathbf{0}_{\text{card}(S) \times (I-1)(J-1)}$, it holds $\mathbf{M}_1(\boldsymbol{\theta}_0, S) \mathbf{M}_2^T(\boldsymbol{\theta}_0, S) = \mathbf{0}_{\text{card}(S) \times \text{card}(S^C)}$ which means that $\mathbf{Z}_1(S)$ and $\mathbf{Z}_2(S)$ are independent, that is

$$\lim_{n \rightarrow \infty} \Pr \left(S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \leq x \right) = \sum_{j=0}^{(I-1)(J-1)} \Pr \left(\chi_{(I-1)(J-1)-j}^2 \leq x \right) w_j(\boldsymbol{\theta}_0)$$

where the expression of $w_j(\boldsymbol{\theta}_0)$ is (21). We have also,

$$\text{Var}(\mathbf{Z}_1(S)) = \mathbf{M}_1(\boldsymbol{\theta}_0, S) \mathbf{M}_1^T(\boldsymbol{\theta}_0, S) = \mathbf{Q}^T(\boldsymbol{\theta}_0, S) \mathcal{I}_F(\boldsymbol{\theta}_0) \mathbf{Q}(\boldsymbol{\theta}_0, S) = \left(\mathbf{R}(S) \mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) \mathbf{R}^T(S) \right)^{-1} = \boldsymbol{\Sigma}_1(\boldsymbol{\theta}_0, S),$$

$$\begin{aligned}
\text{Var}(\mathbf{Z}_2(S)) &= \mathbf{M}_2(\boldsymbol{\theta}_0, S) \mathbf{M}_2^T(\boldsymbol{\theta}_0, S) = \mathbf{R}(S^C) \mathbf{P}(\boldsymbol{\theta}_0, S) \mathcal{I}_F(\boldsymbol{\theta}_0) \mathbf{P}^T(\boldsymbol{\theta}_0, S) \mathbf{R}^T(S^C) \\
&= \mathbf{R}(S^C) \mathbf{P}(\boldsymbol{\theta}_0, S) \mathbf{R}^T(S^C) = \boldsymbol{\Sigma}_2(\boldsymbol{\theta}_0, S).
\end{aligned}$$

The proof of $T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ is almost immediate from the proof for $S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ and taking into account that for some $S \in \mathcal{F}(E)$

$$T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}(S)), \mathbf{p}(\hat{\boldsymbol{\theta}})) + o_p(1) = S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})).$$